

Humboldt-Universität zu Berlin
Philosophische Fakultät I
Institut für Bibliotheks- und Informationswissenschaft

Masterarbeit

Enhanced publications – Integration von Forschungsdaten beim wissenschaftlichen Publizieren

Arbeit zur Erlangung des akademischen Grades Master of Arts

Elena Simukovic

Matrikel-Nr. 542015

Gutachter: 1. Prof. Dr. Peter Schirmbacher
2. Maxi Kindling, M.A.

Abgabedatum: 6. Oktober 2012

Zusammenfassung

In der vorliegenden Arbeit wird die Problematik der Forschungsdatenintegration beim wissenschaftlichen Publizieren analysiert. Dabei werden die disziplinspezifischen Anforderungen hervorgehoben, die sich aus den organisationellen, technischen und politischen Gegebenheiten in den Wissenschaftsdisziplinen ergeben. Zu diesem Zweck wurden fünf Disziplinen ausgewählt, die sich durch unterschiedliche Praktiken im wissenschaftlichen Arbeiten auszeichnen und dadurch eine breite Grundlage für eine Untersuchung bieten. Daneben werden bereits existierende *enhanced publications* beziehungsweise andere Formen von Forschungsdatenintegration in der jeweiligen Disziplin exemplarisch dargestellt. Abschließend erfolgt eine Gegenüberstellung der Entwicklungen in den einzelnen Bereichen hinsichtlich deren Gemeinsamkeiten und Unterschieden.

Schlagwörter:

Forschungsdaten, wissenschaftliches Publizieren, Wissenschaftsdisziplin, gute wissenschaftliche Praxis

Abstract

The purpose of this work was to discuss the issues on integrating underlying research data into scholarly publishing. More specifically, the approach combined case studies of five scientific disciplines with a survey of literature relevant to research data management and special organizational, as well as technical and political aspects in the selected fields. The existing *enhanced publications* and other manifestations of integrated research data were exemplified, respectively. In conclusion, common features and disciplinary differences are given.

Keywords:

Research data, scholarly publishing, scientific discipline, responsible conduct of research

Inhaltsverzeichnis

1. Einleitung	1
1.1. Hintergrund und Motivation	1
1.2. Gegenstand der Arbeit	3
1.3. Vorgehensweise	5
2. Begriffsbestimmung und aktueller Forschungsstand	7
2.1. Definition und Abgrenzung zu verwandten Begriffen	7
2.2. Charakteristische Projekte und Initiativen	10
2.2.1. Zur Zugänglichkeit der Forschungsdaten	10
2.2.2. Zur Vernetzung von Artikel und Daten	15
2.2.3. Zu disziplinspezifischen Anforderungen	17
3. Fallstudien zur Integration von Forschungsdaten	19
3.1. Archäologie	19
3.1.1. Organisation und Arbeitsweise der Disziplin	20
3.1.2. Technische Infrastruktur	21
3.1.3. Politische Rahmenbedingungen	24
3.2. Klimaforschung	26
3.2.1. Organisation und Arbeitsweise der Disziplin	26
3.2.2. Technische Infrastruktur	29
3.2.3. Politische Rahmenbedingungen	32
3.3. Kristallographie	34
3.3.1. Organisation und Arbeitsweise der Disziplin	34
3.3.2. Technische Infrastruktur	36
3.3.3. Politische Rahmenbedingungen	39
3.4. Ökologie	41
3.4.1. Organisation und Arbeitsweise der Disziplin	41
3.4.2. Technische Infrastruktur	43
3.4.3. Politische Rahmenbedingungen	45
3.5. Sozialwissenschaften	47
3.5.1. Organisation und Arbeitsweise der Disziplin	47
3.5.2. Technische Infrastruktur	51
3.5.3. Politische Rahmenbedingungen	54
4. Fazit	56
5. Literaturverzeichnis	59
6. Tabellenverzeichnis	64

7. Abbildungsverzeichnis	65
8. Abkürzungsverzeichnis	66

1. Einleitung

1.1. Hintergrund und Motivation

Forschungsdaten stellen einen unmittelbaren Bestandteil des wissenschaftlichen Arbeitens dar und sind für die umfassende Vermittlung der Forschungsergebnisse (oft bezeichnet als *scholarly record* oder *record of science*) unabdingbar.¹ Die Forderung nach dem Zugang zu den zugrundeliegenden Forschungsdaten bei wissenschaftlichen Publikationen stützt sich vor allem auf die Möglichkeit, die präsentierten Aussagen zu unterstützen. Außerdem soll damit mehr Transparenz bei der Darlegung wissenschaftlicher Erkenntnisse erreicht werden.²

Deshalb verwundert es auch nicht, dass die öffentliche Zugänglichmachung der Forschungsergebnisse einschließlich der Forschungsdaten mit den Grundprinzipien der guten wissenschaftlichen Praxis verbunden wird. Die Erwägungen dazu sind unter anderem in den Richtlinien der nationalen Förderorganisationen zu finden, wie beispielsweise den „Vorschlägen zur Sicherung guter wissenschaftlicher Praxis“³ der Deutschen Forschungsgemeinschaft (DFG) in Deutschland, dem „Australian Code of Responsible Conduct“⁴ des National Health and Medical Research Council (NHMRC) in Australien oder dem „Code of Practice for Research“⁵ des UK Research Integrity Office (UKRIO) in Großbritannien. In den politischen Stellungnahmen der Organisationen wird die Notwendigkeit der Sicherung der Forschungsergebnisse hervorgehoben, damit im Fall des Verdachts auf das wissenschaftliche Fehlverhalten die präsentierten Ergebnisse überprüft bzw. verteidigt werden können.

Fälschungen wissenschaftlicher Forschungsergebnisse stellen in diesem Kontext keine hypothetische Situation dar, sondern treten – bei der technisch möglichen weltweiten Vernetzung der Wissenschaft – vor allem in den letzten Jahren in das öffentliche Interesse. So hatten z.B. die Aufklärung des niederländischen Sozialpsychologen Stapel⁶ im Jahr 2011 oder des koreanischen Zellbiologen Hwang⁷ im Jahr 2006 eine hohe Resonanz. In beiden Fällen ging es um massive Daten-Fälschungen für mehrere Publikationen in wissenschaftlichen Zeitschriften.⁸ Solche und ähnliche Fälle⁹ haben

1 Vgl. Murray-Rust (2007)

2 Vgl. DINI (2009), S. 5

3 http://www.dfg.de/foerderung/rechtliche_rahmenbedingungen/gwp/

4 <http://www.nhmrc.gov.au/guidelines/publications/r39>

5 <http://www.ukrio.org/what-we-do/code-of-practice-for-research/>

6 Vgl. Callaway (2011), <http://dx.doi.org/10.1038/479015a>

7 Vgl. Cyranoski (2009), <http://dx.doi.org/10.1038/4611035a>

8 Vgl. <http://www.tilburguniversity.edu/nl/nieuws-en-agenda/commissie-levelt/interim-report.pdf>

9 Vgl. Editorial (20 September 2012), in: Nature 489 (7416) <http://dx.doi.org/10.1038/489335a>

weitreichende Diskussionen um die Nachprüfbarkeit der Forschungsergebnisse durch die Wissenschaftspolitik angestoßen.¹⁰

Auch die für viele deutsche Wissenschaftsorganisationen als Grundsätze dienenden Empfehlungen der DFG¹¹ wurden durch einen „besonders schwerwiegenden Fall des wissenschaftlichen Fehlverhaltens“ veranlasst.¹² Der freie Zugang zu den die Publikation unterstützenden Forschungsdaten im Zusammenhang mit der Methodenbeschreibung soll deshalb die Möglichkeiten zur Selbstreflexion der Wissenschaft deutlich steigern. Die Verfügbarkeit beider Elemente wird daneben oft als eine Voraussetzung für Reproduzierbarkeit der Forschungsergebnisse (eng. *reproducible research*) angesehen.¹³

Ein weiterer mit der Einbindung von Forschungsdaten beim wissenschaftlichen Publizieren verbundener Aspekt ist das Potential, welches in deren Nachnutzung liegt. Nach den Ergebnissen einer Forscher-Umfrage im PARSE.Insight-Projekt bestätigte die Mehrheit der Teilnehmer, Schwierigkeiten dabei gehabt zu haben, auf die benötigten Forschungsdaten anderer Forscher zuzugreifen.¹⁴ Gleichzeitig beweisen bibliometrische Untersuchungen höhere Zitationszahlen für die mit Forschungsdaten „angereicherten“ Publikationen und damit das große Interesse seitens der Wissenschaft.¹⁵ Auf der politischen Ebene in Europa werden die Potentiale der öffentlich zugänglichen Forschungsdaten über die Fachöffentlichkeit hinaus auf Vorteile für die Gesellschaft und Industrie ausgeweitet. Im Rahmen des EU-Förderprogramms „Horizon 2020“ für die Jahre 2014 bis 2020 werden verschiedene Ansätze verfolgt, um die Zugänglichkeit und Nachnutzung der Daten zu steigern.¹⁶

Die politische Diskussion um Forschungsdaten wird durch den Aspekt der öffentlichen Finanzierung abgerundet. So folgen nationale und internationale Förderorganisationen einem Prinzip, nach dem durch öffentliche Fördermittel entstandene Forschungsergebnisse grundsätzlich öffentlich zugänglich sein sollen.¹⁷ Von der Kommission Zukunft der Informationsinfrastruktur (KII) wird sogar empfohlen, die Forschungsdaten als nationales Kulturgut anzusehen und zukünftigen Generationen zur Verfügung

10 Vgl. „Publishing data in a reusable form to support findings must be mandatory“, in: Science as an open enterprise, The Royal Society Science Policy Centre report 02/12.

<http://royalsociety.org/policy/projects/science-public-enterprise/report/>

11 Vgl. Allianz der deutschen Wissenschaftsorganisationen (2010)

12 DFG (1998), S. 5

13 Vgl. Bechhofer et al. (2011), S. 2

14 Kuipers; van der Hoeven (2009): PARSE.Insight Deliverable D3.4, S. 35

15 Vgl. Piwowar et al. (2007) Daten aus: Sharing Detailed Research Data Is Associated with Increased Citation Rate. PLoS ONE 2(3). Figure 1: <http://dx.doi.org/10.1371/journal.pone.0000308.g001>, Figure 2 <http://dx.doi.org/10.1371/journal.pone.0000308.g002>

16 Europäische Kommission: Scientific data: open access to research results will boost Europe's innovation capacity, Pressemitteilung (IP/12/790) vom 17.07.2012.

17 Vgl. Research Councils UK (2011): RCUK Common Principles on Data Policy <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>

zu stellen.¹⁸ Durch die Nachnutzung der Forschungsergebnisse können außerdem interdisziplinäre Kooperationen entstehen, die neue Fragestellungen beantworten werden.¹⁹

Eine ganz andere Herangehensweise zum Thema bietet die Diskussion um die Entwicklung der wissenschaftlichen Publikation. Kritisiert wird vor allem, dass seit der Gründung der ersten wissenschaftlichen Zeitschriften die Struktur des wissenschaftlichen Artikels sich nicht wesentlich geändert hätte. Der Übergang auf überwiegend digitale Publikationen hat zwar die Geschwindigkeit der Kommunikation stark beschleunigt, die Präsentationsform der Forschungsergebnisse bleibt aber trotzdem an eine „Papier-Logik“ gebunden.²⁰ Neue technische Möglichkeiten wie semantische Verknüpfungen, *text mining* oder *Linked Open Data*, die sich durch die globale Vernetzung der Ressourcen ergeben, können in manchen Fachbereichen bereits eine weite Verbreitung aufweisen.²¹ Zwar stellt das Internet einen wichtigen Katalysator in der Wissenschaftskommunikation dar, doch spielen gerade die disziplinspezifischen Unterschiede im Spannungsfeld zwischen der *data sharing*-Kultur und der den Wissenschaftsbetrieb versorgenden Infrastrukturen eine entscheidende Rolle für die Umsetzung innovativer Methoden.²²

1.2. Gegenstand der Arbeit

In der vorliegenden Masterarbeit wird auf ein Publikationsmodell abgestellt, in dem gängige wissenschaftliche Publikationen in Form eines Artikels in einer Fachzeitschrift mit dazugehörigen, zugrundeliegenden, weitere Forschung unterstützenden Forschungsdaten angereichert werden. Die genauere Definition von Forschungsdaten erfolgt abhängig von den für die Untersuchung ausgewählten Wissenschaftsdisziplinen und wird in den Kapiteln 3 bis 7 exemplarisch dargestellt. Im Allgemeinen bezieht sich dabei der Begriff von Forschungsdaten auf alle im Forschungszyklus entstehenden, sowie zusätzliche wissenschaftliche Publikationen unterstützende Daten inklusive Beobachtungsdaten, Aggregatdaten, Kalkulationen, Visualisierungen, Programmskripten und Methodenbeschreibungen. Aufgrund der besseren Vernetzungsmöglichkeiten bei gleichzeitig niedrigerem Zeit- und Kostenaufwand wird die Analyse auf das elektronische Publizieren beschränkt.

Der Gegenstand dieser Arbeit ist eine „erweiterte“ Form des wissenschaftlichen Zeitschriftenartikels, mit dessen Hilfe die Sichtbarkeit und Auffindbarkeit der zugrundeliegenden Forschungsdaten deutlich erhöht werden kann. Aus diesem Grund wird absichtlich nicht auf private, für kommerzielle Nutzung in der Industrie, zur Patentanmeldung benötigte, aus politischen oder anderen Gründen sensible Daten

¹⁸ KII (2011), S. 29

¹⁹ Ebd., S. 17

²⁰ Vgl. Harnad (1992); Murray-Rust (2007); Lynch (2007); Bechhofer et al. (2011)

²¹ S. Untersuchung der einzelnen Wissenschaftsdisziplinen in folgenden Kapitel

²² Vgl. Meier zu Verl; Horstmann (2011). Introduction, S. 5 ff. <http://dx.doi.org/10.2390/PUB-2011-2>

eingegangen, bei denen eine breite öffentliche Zugänglichkeit nicht angestrebt wird. Als eine optimale technische Lösung wird die Verlinkung zwischen dem Zeitschriftenartikel und den Datensätzen in fachspezifischen Forschungsdaten-Repositoryn angesehen.

Als eine Bezeichnung für den obigen Sachverhalt wird der Begriff von *enhanced publications* übernommen. Zudem werden im Kapitel 2 verwandte Begriffe abgegrenzt, die sich durch eine ähnliche Intention, aber andere Merkmale auszeichnen, so beispielweise die Begriffe *learned article*, *scientific publication packages*, *compound objects*, *executable paper*, *supplemental online material*. Des Weiteren wird auch analysiert, ob und wenn ja, welche Beispiele von *enhanced publications* in ausgewählten Wissenschaftsdisziplinen bereits existieren und welche Voraussetzungen sie erfüllen.

Den Kern der Arbeit bildet das Kapitel 3, in dem die Beantwortung der eigentlichen Forschungsfrage stattfindet. So wird untersucht, ***wie sich die Bedingungen für die Umsetzung von enhanced publications in verschiedenen Wissenschaftsdisziplinen unterscheiden***. Es wurden dafür fünf Wissenschaftsdisziplinen ausgewählt, die sich durch unterschiedliche Praktiken im wissenschaftlichen Arbeiten auszeichnen und dadurch eine breite Grundlage für eine ausführliche Untersuchung bieten:

1. Archäologie,
2. Klimaforschung,
3. Kristallographie,
4. Ökologie,
5. Sozialwissenschaften.

Für die Untersuchung der Wissenschaftsdisziplinen wurde ein Kriterienkatalog erstellt, der auf der aktuellen Literatur zu den fachspezifischen Anforderungen an Publikationsinfrastrukturen und zum Forschungsdatenmanagement basiert. Damit Rückschlüsse für nachhaltige Forschungsdatenintegration gezogen werden können, wird jede Disziplin als ein Ökosystem angesehen, in dem organisationelle, technische und politische Faktoren interagieren.

Im ersten Punkt des Kriterienkatalogs wird zunächst die jeweilige Fachdisziplin vorgestellt, indem die verwendeten Forschungsdaten im Forschungsbereich, typische Arbeitsabläufe und etablierte interne Standards beschrieben werden. Insbesondere wird auf die Charakteristika der Forschungsdaten, zentrale Koordinationseinrichtungen und kennzeichnende Merkmale der Disziplin eingegangen.

Im zweiten Punkt des Kriterienkatalogs wird die technische Infrastruktur der ausgewählten Wissenschaftsdisziplinen untersucht. Dazu werden Funktionalitäten von vorhandenen fachspezifischen Repositoryn oder Daten-Archiven ausgewertet. In Betracht gezogen werden dabei

für *enhanced publications* kritische Aspekte wie Zugänglichkeit von Forschungsdaten, deren persistente Identifizierung und die Beschreibung der Methoden zur Datenerhebung. Alle drei Elemente werden als notwendig für die Nachnutzung und Nachprüfbarkeit der Forschungsdaten angesehen. Des Weiteren werden auch „motivierende“ Aspekte untersucht: direkte Einbindung eines Zitationsvorschlages für Forschungsdaten, Hinweise auf verwandte Publikationen, die aus den Forschungsdaten entstanden sind oder auf diese Bezug haben und Nutzungsstatistiken zu den Download-Zahlen des beschriebenen Datensatzes.

Der Kern von *enhanced publications* liegt in der Integration vom wissenschaftlichen Zeitschriftenartikel und ihm zugrundeliegenden Forschungsdaten. Es wird daher absichtlich nicht auf alle Funktionalitäten von Forschungsdaten-Repositorien eingegangen, sondern nur auf diejenigen Aspekte, die für die Einbindung der Daten beim wissenschaftlichen Publizieren notwendig sind.

Im dritten Punkt des Kriterienkatalogs werden die politischen Rahmenbedingungen untersucht. Dafür werden die verpflichtenden Vorgaben der Förderorganisationen in der Disziplin ausgewertet. Darunter fallen nur Anforderungen zur Projektförderung (eng. *grants*), nicht von einzelnen Forschern (eng. *fellowship, scholarship*).

1.3. Vorgehensweise

Die Integration von Forschungsdaten beim wissenschaftlichen Publizieren soll in einzelnen Analyseschritten untersucht und bewertet werden. Zunächst werden in einem breit angelegten Kapitel verwandte Begriffe durch kennzeichnende Merkmale abgegrenzt. Des Weiteren werden charakteristische Projekte und Initiativen vorgestellt, die maßgeblichen Einfluss auf öffentliche Zugänglichkeit von Forschungsdaten haben und hatten. Daran schließt sich die Vorstellung der Studien zur Verknüpfung von Forschungsdaten und wissenschaftlichen Publikationen sowie zu disziplinspezifischen Unterschieden im Umgang mit Forschungsdaten.

Um die relevanten und bekannten Forschungsdatenrepositorien in der jeweiligen Disziplin zu identifizieren, wurden vier Referenzlisten²³ abgeglichen:

- a) in der Directory of Open Access Repositories registrierte Repositorien²⁴ enthaltend „Datasets“ als „Content type“;
- b) das Arbeitsdokument²⁵ von DataCite in Zusammenarbeit mit der British Library, BioMed Central und dem Digital Curation Centre;

²³ Es wurden einige weitere Register berücksichtigt, die aber als nicht praktikabel abgelehnt werden mussten, darunter Registry of Open Access Repositories (ROAR) und DRIVER Repositorien-Verzeichnis

²⁴ <http://www.openoar.org/find.php> [Stand 20.06.2012]

- c) die vom Open Access Directory zusammengestellte Datenrepositorien-Liste²⁶, betreut von der Graduate School of Library and Information Science am Simmons College;
- d) die Databib-Bibliographie²⁷ der Forschungsdatenrepositorien.

Es wurde eine Mindestzahl von fünf repräsentativen Repositorien je Disziplin angestrebt. Diese sollten zudem in mindestens zwei Referenzlisten verzeichnet werden. Die Ergebnisse der Auswertung werden für die jeweilige Disziplin in einer Tabelle zusammengefasst. Anschließend erfolgt die exemplarische Darstellung der bereits existierenden *enhanced publications* bzw. anderer Formen von Forschungsdatenintegration in der Disziplin. Für die Untersuchung der disziplinspezifischen Vorgaben wurde eine Mindestzahl von drei repräsentativen Förderorganisationen angestrebt. Allerdings fungieren diese meist als national begrenzte Körperschaften, die gleichzeitig mehrere Disziplinen fördern. Deshalb werden auch die allgemeinen Richtlinien verschiedener nationaler Organisationen mit Bezug auf die jeweilige Disziplin berücksichtigt. Abschließend erfolgt eine Gegenüberstellung der Entwicklungen in den einzelnen Bereichen hinsichtlich deren Gemeinsamkeiten und Unterschieden.

25 <http://datacite.org/repolist> [Stand 20.06.2012]

26 http://oad.simmons.edu/oadwiki/Data_repositories [Stand 20.06.2012]

27 <http://databib.org/> [Stand 06.08.2012]

2. Begriffsbestimmung und aktueller Forschungsstand

2.1. Definition und Abgrenzung zu verwandten Begriffen

Die relativ junge Diskussion um *enhanced publications* bietet noch keine feste Definitionen an. Im Laufe der Zeit änderten sich zwar die Fragestellungen bei der Implementierung verschiedener Konzepte, die grundlegende Idee dahinter blieb jedoch konstant: die Vernetzung zusammenhängender Komponenten eines *scholarly records*. Im Kontext der Diskussion um das wissenschaftliche Publizieren werden im vorliegenden Kapitel die relevanten Begriffe voneinander abgegrenzt und das Aufkommen von *enhanced publications* dargestellt.

Die zukünftige Form der elektronischen wissenschaftlichen Publikation wurde bereits in den frühen 1990er Jahren von Fachgesellschaften wie der Coalition for Networked Information (CNI) und der American Physical Society (APS) hinterfragt.²⁸ Dabei wurden mögliche Zukunftsszenarien für akademische Kommunikation im Jahr 2020 diskutiert. Auch wenn die kommenden Entwicklungen sich nicht exakt vorhersehen ließen, waren sich viele Autoren einig, dass die elektronische Publikation viel mehr Interaktionsmöglichkeiten bieten würde als die herkömmlichen Druckexemplare. Daneben verwies Charles W. Bailey auf die Verbesserung der Funktionalität elektronischer Dokumente und die Vernetzung heterogener Informationssysteme als eine notwendige Bedingung für den Erfolg akademischer Publikationssysteme.²⁹

Die Verlinkung einzelner verwandter Objekte mittels Web-Technologien forderte auch Kircz.³⁰ Dabei plädierte er für einen Übergang vom *linearen* zum *modularen* Artikel, der als eine Aggregation zusammenhängender Informationsmodule einschließlich Datensätzen, Bildern, Simulationen, Audio- und Video-Dateien betrachtet wurde. Die Umsetzung der „Papier-Form“ beim elektronischen Publizieren würde seiner Ansicht nach den Bedürfnissen in der Wissenschaftskommunikation nicht mehr entsprechen können.³¹

Gleichzeitig kristallisierte sich die Idee einer Verbindung zwischen der wissenschaftlichen Publikation mit den dazugehörigen Forschungsdaten. So schlugen Rzepa und Murray-Rust unter dem Begriff *learned article* vor, Dokumente und Daten als Bestandteile eines unmittelbaren Spektrums zu

28 Vgl. Harnad (1992); Bailey (1994)

29 Bailey (1994), S. 14

30 Vgl. Kircz (1998, 2001, 2002)

31 Kircz (2002), S. 28 ff.

betrachten. Dabei sei der Zugriff auf die Daten in manchen Wissenschaftsdisziplinen unabdingbar, um ausgiebigen Gebrauch von dem Artikel zu machen.³²

Mit der immer rasanteren Entwicklung der Web-Technologien und wachsenden Datenmengen in der Wissenschaft (oft referenziert als „Datenflut“, eng. *data deluge*)³³ wurden auch datenintensive Forschung unterstützende Konzepte entwickelt. Der von Hunter geprägte Begriff *scientific publication packages* umfasste bereits Primärdaten, Algorithmen, Software und Textpublikationen.³⁴ Das Konzept diene als ein Vorläufer³⁵ für den danach aufgekommenen Begriff *compound objects*: „*Scientific compound objects encapsulate the various datasets and resources generated or utilized during a scientific experiment or discovery process, within a single compound object, for publishing and exchange*“.³⁶ Eine konkrete Anwendung nach diesen Prinzipien stellt beispielsweise das im Rahmen des eSciDoc-Projekts entwickelte Publikationsmanagement-System der Max-Planck-Gesellschaft dar.³⁷ Die Objekte werden in *containers* sortiert und die Integration verwandter Inhalte durch das Ansetzen von Beziehungstypen ermöglicht.³⁸ Allerdings wird dabei keine besondere Beziehung zwischen Publikationen und Forschungsdaten hervorgehoben (wie etwa Unterstützung oder Nachprüfbarkeit), sondern eher ein Mittel für die Organisation und den Austausch digitaler Objekte zur Verfügung gestellt.³⁹

Der Zusammenhang verschiedener Elemente wird hingegen unter dem Begriff *research objects* angesprochen. Als strukturierte semantische Aggregationen in Linked Data sollen damit Daten, Methoden und Personen verknüpft werden.⁴⁰ Noch mehr strukturorientiert zeigen sich weitere Konzepte der Semantic Web-Gemeinschaft. So bringt beispielsweise Shotton den Begriff des *semantic publishing* ein: „*anything that enhances the meaning of a published journal article, facilitates its automated discovery, enables its linking to semantically related articles, provides access to data within the article in actionable form, or facilitates integration of data between papers*“.⁴¹ Von zentraler Bedeutung ist dabei die automatisierte Vernetzung und maschinelle Weiterverarbeitung der Publikationen.

32 Rzepa; Murray-Rust (2001)

33 Vgl. Blake; Bult (2006), S. 314; Hunter (2006), S. 33; Borgman (2008), S. 30

34 Hunter (2006), S. 42 ff.

35 Vgl. Bechhofer et al. (2011), S. 3

36 Cheung, Hunter et al. (2008), S. 4

37 <http://colab.mpg.de/mediawiki/Portal:ESciDoc>

38 <https://www.escidoc.org/JSPWiki/en/ContentModel>

39 <https://www.escidoc.org/JSPWiki/en/ESciDoc>

40 Bechhofer et al. (2011), S. 11

41 Shotton (2009), S. 86

Einer ähnlichen Intention folgt auch der nach dem „Elsevier Executable Paper Grand Challenge“ sich verbreitete Begriff *executable paper*.⁴² So fügt beispielsweise Kauppinen im Vergleich zu den oben dargestellten Konzepten die Nutzung von Open Source Software und CC-Lizenzen als Merkmal hinzu.⁴³ Allerdings steht bei solchen und vergleichbaren Begriffen oft die pragmatische Optimierung der Auffindbarkeit und der Verlinkung von Internetressourcen im Mittelpunkt. Da aber die Integration von Forschungsdaten beim wissenschaftlichen Publizieren an vielerlei organisatorische, kulturelle und politische Faktoren gebunden ist, wurde in der vorliegenden Masterarbeit nach einem Begriff gesucht, der in seiner Definition auch weitere Aspekte dieser Form der wissenschaftlichen Publikation umfassen würde.

Die wohl allumfassendste Betrachtung des Themas ist derzeit aus dem innerhalb des 7. Rahmenprogramms für Forschung der Europäischen Union (FP7) geförderten Projektes „Digital Repository Infrastructure Vision for European Research II“ (DRIVER II) ersichtlich. Dabei wurde ein Ansatz gewählt, bei dem die bedeutende Rolle von Forschungsdaten im Mittelpunkt steht: „*A promising approach to improving the access to research data is through the enhancement of the traditional publication. This entails, more concretely, that publications are enriched with references to the primary data that were used to produce the insights that are put forward in this text*“.⁴⁴ Über die Unterstützungsfunktion hinaus erweitert Hoogerwerf dabei den Begriff der *enhanced publications* mit dem Aspekt der Nachnutzbarkeit: „*Enhanced publications are compositions of textual publications and supporting resources. In addition to the possibility to support their textual publication with e.g. data or visualizations, they also promote the availability of reusable scientific data and allow verification of the outcomes of the research*“.⁴⁵ Daneben wird noch eine erweiterte Perspektive zur Anreicherung bei *enhanced publications* angeboten: „*We define an enhanced publication as a publication that is enriched with three categories of information: research data (evidence of the research); extra materials (to illustrate or clarify); post-publication data (commentaries, ranking)*“.⁴⁶ So werden gleichzeitig die Nachprüfbarkeit, Nachnutzbarkeit und Rezeption des wissenschaftlichen Artikels unterstützt.

Außerdem werden im vorwiegend von niederländischen Forschungseinrichtungen durchgeführten DRIVER II Projekt auch technische Anforderungen beschrieben. So wird für die Umsetzung von *enhanced publications* ausschließlich das OAI-ORE Modell empfohlen, das insbesondere aufgrund der dynamischen Natur der angereicherten Publikation als geeignet erscheint.⁴⁷ Die Umsetzung unter Anwendung von Resource Maps, Persistent Identifiers und Resource Description Framework (RDF)

⁴² <http://www.executablepapers.com/>

⁴³ Kauppinen (2011), S.726

⁴⁴ Verhaar (2008): DRIVER II Deliverable D4.2, S. 37

⁴⁵ Hoogerwerf (2009), Durable Enhanced Publications (ohne Seitennummer)

⁴⁶ Woutersen-Windhower; Brandsma (2009): DRIVER II Deliverable D4.1, S. 7

⁴⁷ Ebd., S.7; Deliverable D4.2, S. 21

wird von der ebenfalls am Projekt beteiligten niederländischen Dachorganisation SURF hervorgehoben.⁴⁸

In der vorliegenden Masterarbeit werden *enhanced publications* aus einer eher pragmatischen Sicht betrachtet. Als *enhanced publications* sollen alle diejenigen Publikationen angesehen werden, die traditionell in Form eines elektronischen Artikels in einer fachwissenschaftlichen Zeitschrift publiziert werden und die dem Artikel zugrundeliegenden Forschungsdaten einschließen. Als eine optimale technische Umsetzung wird in diesem Kontext die Verlinkung zwischen dem Zeitschriftenartikel und den Forschungsdaten in einem fachspezifischen Repository oder Datenarchiv angesehen.

Zusätzliche unterstützende Materialien werden derzeit von zahlreichen wissenschaftlichen Zeitschriften als *supplemental online material* (SOM) publiziert.⁴⁹ Trotz zunehmender Verbreitung⁵⁰ wird eine solche Vorgehensweise oft aufgrund wesentlicher Nachteile kritisiert. Das Opportunities for Data Exchange (ODE) Projekt verweist dabei auf die fehlenden Standards zur Referenzierbarkeit und dadurch mangelhafte Auffindbarkeit des zusätzlichen Materials.⁵¹ Daneben zeigt das vorwiegend auf Forschungsdaten in den Biowissenschaften ausgerichtete Dryad-Repository die Unterschiede zwischen Diensten der Forschungsdatenrepositorien und der Einreichung als SOM noch detaillierter auf. Hervorgehoben wird dabei insbesondere die eindeutige Identifizierung und professionelle Verwaltung der Daten.⁵² Aufgrund dieser Einschränkungen wird die Integration von Forschungsdaten als SOM im Rahmen der vorliegenden Arbeit nicht näher analysiert.

2.2. Charakteristische Projekte und Initiativen

2.2.1. Zur Zugänglichkeit der Forschungsdaten

Im folgenden Abschnitt werden verschiedene Projekte und Initiativen dargestellt, die einen maßgeblichen Einfluss auf die Zugänglichkeit von Forschungsdaten haben.

Einen Meilenstein im Umgang mit digitalen Forschungsdaten stellt das Human Genome Project (HGP) dar. Das vom U.S. Department of Energy und den National Institutes of Health (NIH) koordinierte Projekt wurde unter der weltweiten Beteiligung von mehreren Forschungseinrichtungen

48 <http://www.surf.nl/en/themas/openonderzoek/verrijktepublicaties/Pages/default.aspx>

49 Vgl. Michel et al. (2011): Quantitative Analysis of Culture Using Millions of Digitized Books, Science 331 (6014) <http://dx.doi.org/10.1126/science.1199644>

50 Vgl. Schriger et al. (2011): Use of the Internet by Print Medical Journals in 2003 to 2009: A Longitudinal Observational Study, Annals of Emergency Medicine (57/2) <http://www.annemergmed.com/article/S0196-0644%2810%2901648-3/fulltext#tbl1>

51 Reilly et al. (2011): ODE Report, S. 37

52 Data archiving: Dryad repository & journal Supplementary Online Materials (SOM), updated June 2012. http://wiki.datadryad.org/wg/dryad/images/b/b0/Dryad_vs_SupMatREVISED2011.06.14.pdf

im Jahr 2003 abgeschlossen.⁵³ In seiner 13-jährigen Laufzeit ist es Wissenschaftlern gelungen, die Gensequenzen der chemischen Basispaare, welche die DNA (Desoxyribonukleinsäure, eng. *deoxyribonucleic acid*) des Menschen bilden, komplett zu identifizieren. Dieses Ergebnis, oft *großzügig* bezeichnet als Entschlüsselung des menschlichen Erbguts, gilt als revolutionär für alle Disziplinen der Lebenswissenschaften.⁵⁴ Das Ausmaß des Projektes zeichnet sich dazu durch eine enge Zusammenarbeit verschiedener Wissenschaftsbereiche aus, die das Projekt erst ermöglicht hat. Diese wird nicht zuletzt vom Logo des HGP repräsentiert (Abbildung 1).



Abbildung 1: Das Logo des Human Genome Project (Quelle: U.S. Department of Energy Genome Programs)

Trotz der Kritik⁵⁵ an dem durch das HGP generierten wirtschaftlichen Wert dient es als eines der wichtigsten Beispiele, wie die Kooperation verschiedener Wissenschaftsdisziplinen und offener Zugang zu den Forschungsdaten neuartige Fragestellungen ermöglichen kann.⁵⁶

Eine vom Publishing Research Consortium (PRC) im Jahr 2010 durchgeführte Studie deutete allerdings auf eine immer noch bestehende Kluft zwischen dem Zugang zu und der Bedeutung von Forschungsdaten für das wissenschaftliche Arbeiten hin.⁵⁷ So stellen die Datensätze, Datenmodelle, Algorithmen und Programme eine der wichtigsten Informationsressourcen dar, die aber nur schwer zugänglich sind (Abbildung 2).

⁵³ http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml

⁵⁴ Vgl. Mardis (2011): A decade's perspective on DNA sequencing technology, *Nature* 470 (7333) <http://dx.doi.org/10.1038/nature09796>

⁵⁵ Vgl. Drake (2011): What is the human genome worth? *Nature News* <http://dx.doi.org/10.1038/news.2011.281>

⁵⁶ Vgl. Giaretta et al. (2010), S. 11

⁵⁷ PRC (2010), S. 9

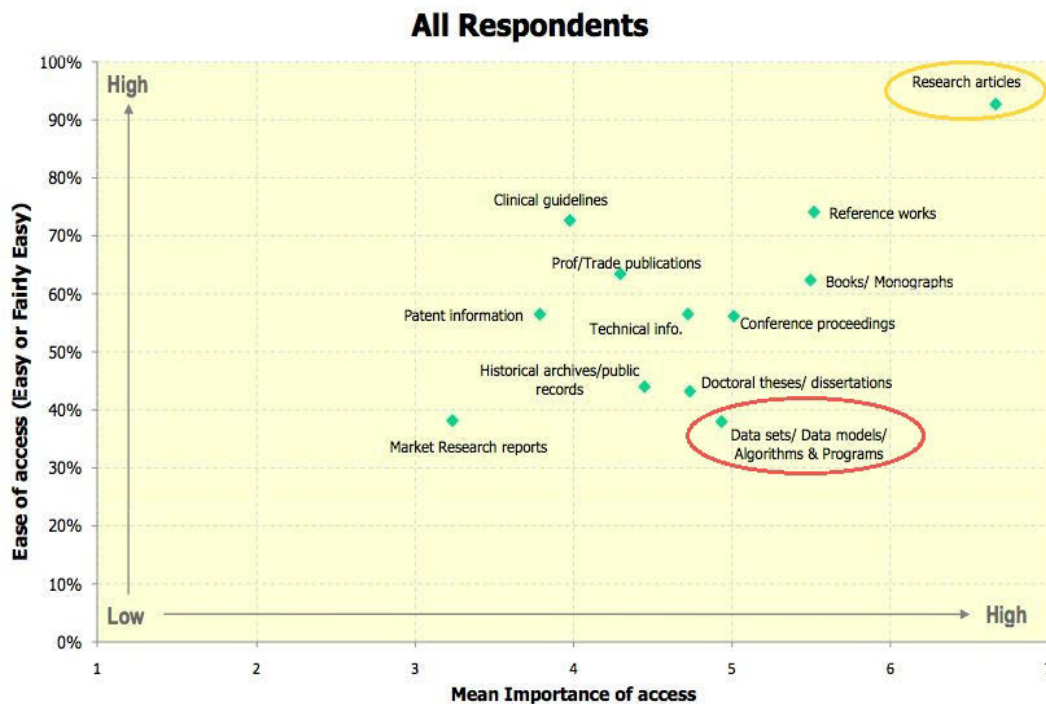


Abbildung 2: Zugang zu vs. Bedeutung von Informationsressourcen
(in Anlehnung an PRC (2010), S. 9)

Während die Wertschätzung und die Zugangsmöglichkeiten zu den als Zeitschriftenartikel publizierten Forschungsergebnissen auf eine fest etablierte Lage hinweisen, signalisiert die Positionierung der Forschungsdaten noch zu überwindende Barrieren vor allem in deren Zugänglichkeit. Für Umsetzung von *enhanced publications* ist dabei die Kombination beider Elemente – Forschungsdaten und Artikel – entscheidend. Damit wird auch dem Standpunkt im DRIVER II-Projekt Rechnung getragen, wonach der Zugang zu Forschungsdaten am besten über die Verbindung mit dem traditionell publizierten wissenschaftlichen Artikel verbessert werden kann.

Die Gründe für den Mangel an Zugänglichmachung von Forschungsdaten (oft bezeichnet als *data sharing*) wurden unter anderem im „Permanent Access to the Records of Science in Europe“ (PARSE) Projekt untersucht. So gaben die meisten Forscher rechtliche Bedenken und den möglichen Missbrauch ihrer Daten als hauptsächliche Barriere für das *data sharing* an.⁵⁸ Als eine gute Aussicht für die Entwicklung von *enhanced publications* kann in der Untersuchung allerdings die Erkenntnis gewertet werden, dass die Nachnutzung, Nachprüfbarkeit und die Unterstützung der wissenschaftlichen Forschungsergebnisse als die drei wichtigsten Motivationsfaktoren für die Langzeiterhaltung und Bereitstellung von Forschungsdaten angesehen werden.⁵⁹

⁵⁸ Kuipers; van der Hoeven (2009): PARSE.Insight Deliverable D3.4, S. 34 (Figure 20)

⁵⁹ Ebd., S. 23

Konkrete Fortschritte für eine bessere Referenzierbarkeit und Zitierbarkeit der Forschungsdaten werden mit dem Einsatz von sog. persistenten Identifikatoren (PID) gemacht. Diese haben sich im Bereich des wissenschaftlichen Publizierens vor allem für die langfristige Kennung digitaler Objekte bewährt. So sind es beispielsweise Uniform Resource Name (URN), persistent URL (PURL), Archival Resource Key (ARK) und Digital Object Identifier (DOI).⁶⁰ Der einem Objekt zugeteilte Identifikator bleibt in dem Sinne persistent, dass er sich nicht ändert, aber durch das interne Resolving-System immer zu der aktuellen URL des Objektes aufgelöst wird. Ein solcher Vorgang stellt den geringsten Aufwand für den Nutzer dar und erhöht gleichzeitig die Auffindbarkeit der auf diese Weise referenzierten Ressourcen.⁶¹

Die Auflösung von DOI bietet manche Vorteile gegenüber anderen PID-Systemen. So sind z.B. die in Deutschland und den Niederlanden oft eingesetzten Uniform Resource Names for National Bibliography Numbers (URN:NBN)⁶² nur durch die Angabe eines nationalen URN:NBN-Resolvers aufzulösen.⁶³ Im Vergleich dazu basiert der DOI auf einem globalen Handle-System, das es ermöglicht, unabhängig von der Registrierungsagentur einen DOI durch die Eingabe des Proxy „<http://dx.doi.org/>“ aufzulösen.⁶⁴ Daneben stellt gerade der DOI einen de facto-Standard beim elektronischen Publizieren dar.⁶⁵ Von besonderer Bedeutung für *enhanced publications* ist die Vergabe von DOI-Namen für Forschungsdaten durch das im Jahr 2009 gegründete Konsortium DataCite.⁶⁶ Mit über 1,3 Millionen registrierten DOIs ist das DataCite eine treibende Kraft für eine höhere Wertschätzung und effizientere Einbindung von Forschungsdaten in wissenschaftlichen Publikationen.⁶⁷

Die ausschlaggebende Rolle des DOI-Systems wird durch die jüngsten Entwicklungen noch deutlicher. So wurde es im Mai 2012 durch die International Organization for Standardization (ISO) zum ISO-Standard 26324:2012 zur Dokumentation digitaler Objekte erklärt.⁶⁸ Außerdem folgte im Juni 2012 die Unterzeichnung eines Joint Statement zwischen DataCite und der International Association of Scientific, Technical & Medical Publishers (STM). Die Verständigung beider Organisationen umfasst gleichzeitig die wichtigsten *Stakeholder*-Gruppen bezüglich der Verlinkung

60 Vgl. Bellini (2012): APARSEN Deliverable D22.1, S. 9 ff.

61 Vgl. DOI Handbook, Chapter 1 – Introduction http://www.doi.org/doi_handbook/1_Introduction.html

62 Vgl. <http://persid.org>

63 <http://www.persistent-identifier.nl/> für Niederlande und <http://nbn-resolving.org/> für Deutschland, Österreich und Schweiz

64 http://www.doi.org/doi_handbook/2_Numbering.html#2.6.2

65 Vgl. <http://www.info.sciverse.com/sciencedirect/implementation/linking/CrossRef>

66 Vgl. <http://www.tib-hannover.de/de/dienstleistungen/doi-service/>

67 Präsentation von Jan Brase am Workshop zum Thema „Metadata and Persistent Identifiers for Social Science Data“, S. 5. Stand Mai 2012 http://www.ratswd.de/ver/docs_PID_2012/Brase_PID2012.pdf

68 ISO Pressemitteilung vom 10.05.2012 <http://www.iso.org/iso/news.htm?refid=Ref1561>

zwischen Zeitschriftenartikeln und zugrundeliegenden Forschungsdaten: die Archive und Bibliotheken seitens DataCite mit den wissenschaftlichen Verlagen seitens STM.⁶⁹

Des Weiteren hat sich dem Dokument im August 2012 auch die überwiegend auf DOI-Vergabe für wissenschaftliche Artikel ausgerichtete CrossRef-Organisation angeschlossen.⁷⁰ Als zwei DOI-Registrierungsagenturen kooperieren CrossRef und DataCite bereits in einigen Initiativen zur Verbesserung der Wissenschaftskommunikation wie z.B. zur Entwicklung eines „DOI Content Negotiation“-Service. Dieser Service soll künftig ermöglichen, die Metadaten von durch CrossRef registrierten Publikationen und von durch DataCite registrierten Forschungsdaten automatisch auszutauschen.⁷¹ Nach dem Abschluss der Beta-Phase soll dadurch eine Art von automatisierten *enhanced publications* entstehen.

Zu den jüngsten Initiativen zur Zitierbarkeit von Forschungsdaten zählt der von Thomson Reuters in Partnerschaft mit dem Inter-university Consortium for Political and Social Research (ICPSR)⁷² bereits zum Ende des Jahres 2012 angekündigte *Data Citation Index*.⁷³ Mit seinem „Web of Science“-Angebot stellt Thomson Reuters einen der wichtigsten Akteure für quantitative Indikatoren der Wissenschaftskommunikation dar. Gerade der nach der Indexierung dieser Datenbank berechnete *Journal Impact Factor* (JIF) ist zur Zeit der wichtigste Indikator, nach dem das Ansehen und Prestige einer Zeitschrift berechnet wird.⁷⁴ Seine Auswirkungen hat dieser Indikator oft direkt in der Belohnungsstruktur, indem Publizieren der Beiträge in den Zeitschriften mit dem bestimmten JIF für die Besetzung von mancher Forschungspositionen vorausgesetzt werden.⁷⁵ Gerade durch Zitation, die oft als Währung der Wissenschaft bezeichnet wird, wird die Leistung des Autors einer Publikation anerkannt. Der *Data Citation Index* soll nach dieser Logik die Motivation der Wissenschaftler steigern, die Forschungsdaten geeignet zu publizieren.

69 http://www.stm-assoc.org/2012_06_14_STM_DataCite_Joint_Statement.pdf

70 http://www.stm-assoc.org/2012_08_10_CrossRef_Joins_STM_DataCite_Statement.pdf

71 DOI Content Negotiation-Beta <http://data.datacite.org/static/> [Stand 13.08.2012]

72 ICPSR Pressemitteilung vom 22. Juni 2012

<http://www.icpsr.umich.edu/icpsrweb/ICPSR/support/announcements/2012/06/data-citation-index-from-thomson>

73 Thomson Reuters Pressemitteilung vom 22. Juni 2012 ;

http://thomsonreuters.com/content/press_room/science/686112

74 Vgl. Garfield (2005), S. 18

75 Vgl. Bewerbungshinweise für Professuren an der Charité-Universitätsmedizin Berlin, Punkt V. Publikationstätigkeit: „vollständige Publikationsliste, darunter Originalarbeiten als Erstautor/in, Seniorautor/in, Impact-Punkte insgesamt und in den letzten 5 Jahren, darunter jeweils gesondert ausgewiesen als Erst- und Seniorautor/in, persönlicher Scientific Citations Index (SCI, h-Index nach Web of Science) über alle Arbeiten“ (nach B.Brembs bei der SWIB11-Konferenz);

http://www.charite.de/charite/organisation/karriere/stellenboerse/bewerbungshinweise_fuer_professuren/

Schließlich ist eine zentrale Suchstelle der für langfristige Sicherung und Referenzierbarkeit der Forschungsdaten geeigneten Repositorien von hoher Bedeutung. Dieser Aufgabe haben sich zwei junge Initiativen - Databib⁷⁶ und Re³Data⁷⁷ - gewidmet. Mit dem globalen Register von Forschungsdaten-Repositorien soll eine zentrale Anlaufstelle für alle beteiligten Interessengruppen geschaffen werden. Die Kombination verschiedener Initiativen kann dabei der Akzeptanz von Forschungsdaten als einem gleichwertigen Bestandteil des *scientific record* erheblich beitragen.⁷⁸

2.2.2. Zur Vernetzung von Artikel und Daten

Anschließend an die Übersicht der Projekte und Initiativen zur Zugänglichkeit und Zitierbarkeit der Forschungsdaten wird im folgenden Abschnitt auf maßgebliche Projekte eingegangen, die sich gezielt mit der Verknüpfung von Forschungsdaten und Publikationen befassen.

Aggregationen zusammenhängender Ressourcen mittels Semantic Web-Architekturen wurden neben dem DRIVER-II-Projekt⁷⁹ auch im ESCAPE-Projekt⁸⁰ modelliert. Dieses von der niederländischen SURFfoundation koordinierte Projekt verkörperte die Idee von angereicherten wissenschaftlichen Publikationen durch den Einsatz des OAI-ORE Modells.⁸¹ Die Zielsetzung des Projektes wurde als Vernetzung verwandter Objekte im Sinne von *compound objects* gesehen.⁸²

„To meet the above mentioned need for ‘enhanced’ scientific communication, the ESCAPE-project has been initiated in order to extend the existing infrastructure of repositories of scientific publications in such a way that it will be possible to identify, describe, preserve and present aggregations of related objects (documents, videos, datasets, etc.), not necessarily produced by the same individual author or group of authors.“

Der Ansatz stellt eine im Semantic Web typische Herangehensweise dar, indem alle durch eine URI beschreibbaren Objekte in eine semantische Verknüpfung gesetzt werden können. So eine Verbindung aller relevanten Objekte könnte ein sinnvolles Lösungsmodell für das Konzept einer *nano-publication* darstellen, indem thematische *statements* in der wissenschaftlichen Literatur aggregiert werden.⁸³ Wenngleich eine solche Vernetzung der Objekte für die weitere Forschung sicherlich interessant ist,

76 <http://databib.org/about.php>

77 <http://www.re3data.org/about/>

78 Vgl. <http://www.re3data.org/2012/04/memorandum-of-understanding-between-datacite-and-re3data-org/>

79 <http://www.driver-repository.eu/Driver-About/About-DRIVER.html>

80 <http://www.surf.nl/en/projecten/Pages/ESCAPE-Enhanced-Scientific-Communication-by-Aggregated-Publications-Environments.aspx>

81 <http://www.openarchives.org/ore/>

82 van Bentum et al. (2009), S. 2

83 Vgl. Mons; Velterop (2009); Groth, Gibson; Velterop (2010)

wird sich die vorliegende Arbeit aufgrund der Ermöglichung der Nachprüfbarkeit und Nachnutzung der Forschungsergebnisse auf die besondere Unterstützungsfunktion der zugrundeliegenden Forschungsdaten in *enhanced publications* konzentrieren.

Die vielseitigen Aspekte der Integration von Daten und Publikationen werden insbesondere im Opportunities for Data Exchange (ODE)-Projekt⁸⁴ analysiert. Die zum großen Teil auf PARSE-Ergebnissen basierende Studie liefert unter anderem die häufig zitierte „Data Publication Pyramid“, in der mögliche Erscheinungsformen von Daten in wissenschaftlichen Publikationen dargestellt werden (Abbildung 3).

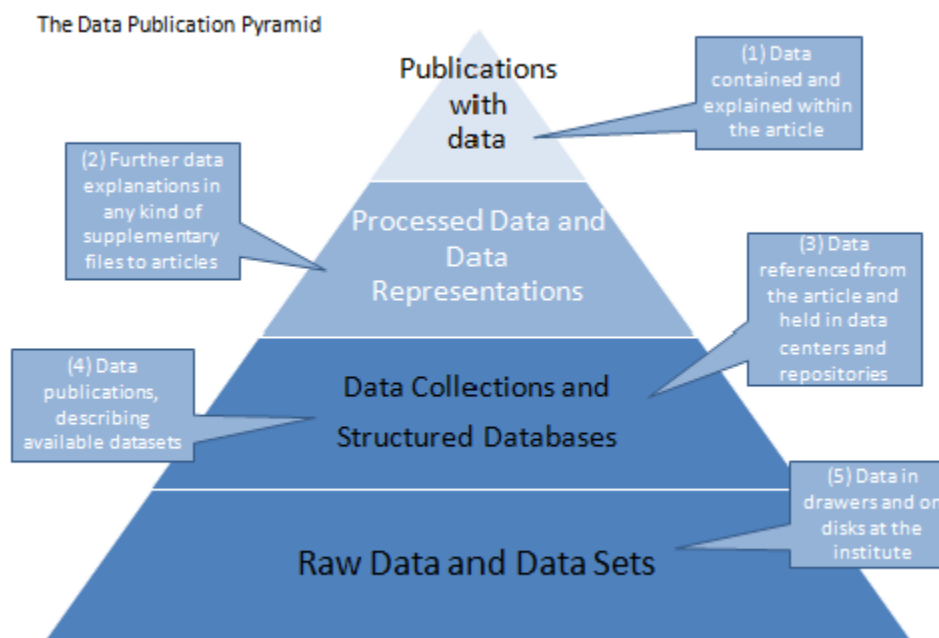


Abbildung 3: Die Datenpublikationspyramide (Quelle: ODE (2011), S. 36)

Den größten Anteil von nicht lokal gespeicherten Daten stellen dabei die beschreibenden Datenpublikationen (sog. *data papers*) und mit den Daten in Repositorien vektürte Publikationen (im Sinne von *enhanced publications*) dar. Als repräsentative Beispiele für solche Formen der Datenintegration genannte Repositorien und Zeitschriften wie PANGAEA, CCDC, PubChem und ESDS⁸⁵ werden in der vorliegenden Arbeit bei der Untersuchung der einzelnen Wissenschaftsdisziplinen näher analysiert.

Auch auf lokaler Universitäts- bzw. Institutions-Ebene sind Projekte zur Integration von Forschungsdaten und Publikationen zu finden. So werden z.B. im LAIRD-Projekt an der University of

⁸⁴ <http://www.alliancepermanentaccess.org/index.php/community/current-projects/ode/>

⁸⁵ Reilly et al. (2011): ODE Report, S. 38 ff.

Edinburgh die Bestände in den zwei eigenen institutionellen Repositorien „Edinburgh DataShare“⁸⁶ (für digitale Forschungsdaten) und „Research Publications Service“⁸⁷ (für Publikationen) gegenseitig verlinkt.⁸⁸ Ähnlich werden im CLADDIER-Projekt die Publikationen im institutionellen Repository der Southampton University mit den umweltwissenschaftlichen Daten im British Atmospheric Data Centre verlinkt.⁸⁹ Ein analoges Beispiel in den Sozialwissenschaften stellt das InFoLiS-Projekt zwischen GESIS – Leibniz-Institut für Sozialwissenschaften und der Universität Mannheim dar.⁹⁰

Ein zusätzlicher Aspekt zu den bereits vorgestellten Ansätzen wird in der zweiten Generation des „Open Access Infrastructure for Research in Europe“ (OpenAIRE) Projektes – dem OpenAIREplus – hervorgehoben.⁹¹ Das im Rahmen des FP7-Programms in den Jahren von 2011 bis 2014 laufende Projekt wird die Infrastruktur des Vorgängers OpenAIRE nutzen, um durch die EU-Förderung entstandene Zeitschriftenartikel und Forschungsdaten sowie Förderinformationen zu verknüpfen.⁹² Damit sollen nicht nur die den Publikationen zugrundeliegende Daten verlinkt werden, sondern auch die *outputs* aus den geförderten Projekten nachvollziehbar und ersichtlich gemacht werden. Auf solche Weise geschaffene Transparenz in der Forschung kann künftig auch für andere Förderorganisationen von Interesse sein und zur Zugänglichkeit der aus öffentlichen Fördermitteln finanzierten Forschungsergebnissen erheblich beitragen.

2.2.3. Zu disziplinspezifischen Anforderungen

Im Anschluss an die Übersicht der Projekte zur Vernetzung von Artikeln und Daten wird im folgenden Abschnitt auf die maßgeblichen Projekte eingegangen, die disziplinspezifische Anforderungen an Informationsinfrastrukturen untersuchten.

Von einer „doppelten“ Bedeutung für die vorliegende Masterarbeit ist das von University of Edinburgh koordinierte „Source to Output Repositories“ (StORe) Projekt.⁹³ Einerseits widmete sich das Projekt dem Zusammenspiel zwischen den Repositorien mit Forschungsprimärdaten (als *source*) und Repositorien mit Publikationen (als *output*).⁹⁴ Andererseits hat es durch seine Vorgehensweise die disziplinspezifischen Anforderungen an beide Repositorientypen aus der Sicht der ausgewählten

⁸⁶ <http://datashare.is.ed.ac.uk/>

⁸⁷ <http://www.ed.ac.uk/schools-departments/information-services/services/research-support/publish-research/research-publications>

⁸⁸ <http://www.ed.ac.uk/schools-departments/information-services/about/organisation/edl/data-library-projects/laird>

⁸⁹ <http://www.jisc.ac.uk/whatwedo/programmes/digitalrepositories2005/claddier>

⁹⁰ <http://www.gesis.org/forschung/drittmittelprojekte/projektuebersicht-drittmittel/infolis/>

⁹¹ <http://www.openaire.eu/en/about-openaire/openaireplusproject/fact-sheet>

⁹² <http://www.openaire.eu/en/support/faq> Nr. 50 und 54

⁹³ <http://www.jisc.ac.uk/whatwedo/programmes/digitalrepositories2005/store.aspx>

⁹⁴ Web-Portale von wissenschaftlichen Verlagen bzw. Zeitschriften wurden dabei auch als Output-Repositorien angesehen

Wissenschaftsdisziplinen analysiert. Die Erkenntnisse aus dem Disziplinen-Vergleich in dem Projekt wurden für die Entwicklung einer Middleware und eines Business-Modells eingesetzt. Wenngleich die Forschungsfragen im StORe-Projekt und in der vorliegenden Masterarbeit sehr nah liegen, war es zu erwarten, dass die Kommunikationskanäle in der Wissenschaft sich in den letzten Jahren stark verändert haben. So wurde z.B. in StORe-Umfrage festgestellt, dass über 50% des *data sharing* über Post- oder E-Mail-Korrespondenz erfolgt.⁹⁵ Aufgrund der wachsenden Datenmengen wäre eine solche Vorgehensweise aktuell nicht mehr praktikabel. Des Weiteren stellt der Kern der Studie – die Forscherperspektive – nur einen Teil der Aspekte für *enhanced publications* dar.

Ein weiteres Projekt, das den Umgang mit Forschungsdaten aus der disziplinspezifischen Sicht untersuchte, wurde vom schottischen Digital Curation Center (DCC) beauftragt. Unter dem Titel „Disciplinary Approaches to Sharing, Curation, Reuse and Preservation“ (SCARP) wurden mehrere Fallstudien in ausgewählten Wissenschaftsdisziplinen durchgeführt.⁹⁶ Im Unterschied zu vergleichbaren Studien dieser Art wurden die Projektmitarbeiter zusätzlich zu Literaturanalyse und Umfragen bei den Fachorganisationen für bis zu 27 Monate eingegliedert.⁹⁷ Dies ermöglichte der Studie auch implizite Arbeitsabläufe im Umgang mit den Forschungsdaten zu erkennen und stellt somit eine wichtige Ergänzung zu den etablierten Praktiken in den Fachdisziplinen dar.

Insbesondere wurden im SCARP organisationelle bzw. kulturelle Faktoren identifiziert, die für die Unterschiede in den Disziplinen verantwortlich sind. Dazu gehören z.B. Einmaligkeit der produzierten Daten, Konkurrenz in dem Fachbereich und Richtlinien der Förderorganisationen.⁹⁸ Diese Erkenntnisse wurden bei der Zusammenstellung des Kriterienkatalogs für die Beantwortung der Forschungsfrage dieser Masterarbeit berücksichtigt. Daneben wurden im Projekt teilweise sehr große Unterschiede zwischen den Forschungsbereichen identifiziert, sodass anstelle der „Wissenschaftsdisziplinen“ die *domains* oder sogar spezialisierte Forschergruppen analysiert werden sollten. Dies wurde bei der Auswahl der zu analysierenden Fachbereiche in der vorliegenden Masterarbeit ebenso berücksichtigt. Des Weiteren wurden die Projektergebnisse in *best practice*-Empfehlungen für verschiedene Interessengruppen umgesetzt, in denen auch die direkte Zitierung von zugrundeliegenden Forschungsdaten sowie deren Archivierung in geeigneten Repositorien empfohlen wird.⁹⁹

95 Pryor (2007): Project StORe: Work Package 5, Evaluation. S. 10

96 <http://www.dcc.ac.uk/projects/scarp>

97 Lyon et al. (2010): DCC SCARP Final Report, S. 9

98 Key Perspectives Ltd. (2010): DCC SCARP Synthesis Report, S. 25 ff.

99 Lyon et al. (2010): DCC SCARP Final Report, S. 20 ff.

3. Fallstudien zur Integration von Forschungsdaten

3.1. Archäologie

3.1.1. Organisation und Arbeitsweise der Disziplin

Forschungsdaten in der Disziplin

Die Archäologie wird oft als eine Wissenschaftsdisziplin der Geistes- und Sozialwissenschaften gesehen (eng. *social sciences and humanities*, auch *arts and humanities*).¹⁰⁰ Dabei betreibt der Fachbereich eine interdisziplinär orientierte Forschung. So werden beispielsweise neben historischen Analysen von Relikten des Altertums Kenntnisse von Materialeigenschaften und Konservierungsverfahren oder der Einsatz von Geoinformationssystemen (GIS) zur Lokalisierung und Dokumentation der Grabungen mit einbezogen. Auch die Forschungsdaten in der Archäologie sind sehr heterogen. Sie umfassen unter anderem Textberichte, Fotografien von Fundstücken und Ausgrabungsstätten, Zeichnungen zur Rekonstruktion von Grabungsobjekten, digitalisierte Pläne und Karten, Satellitenbilder mit Fernerkundungsdaten (eng. *remote sensing*) und topographischen Daten zur Darstellung und Beschreibung der Erdoberfläche.^{101,102}

Eine besondere Art von Forschungsdaten in der Archäologie sind die bei Grabungen aufgefundenen physischen Artefakte¹⁰³ (eng. *artefacts*, bzw. *artifacts*). Diese werden zunächst von den Forschern untersucht und dokumentiert und schließlich konserviert. Die Aufgabe der Langzeitarchivierung wird dabei in der Regel von den „physischen Repositorien“ – dafür zuständigen Museen – übernommen.¹⁰⁴

Der Gegenstand und die Arbeitsweise der Disziplin bedingen somit spezielle Anforderungen an die Erschließung der Forschungsdaten in den Datenarchiven oder Repositorien. Das Metadatenschema des Repository muss gegebenenfalls Angaben zum lokalen Aufbewahrungsort des zu beschreibenden Objektes und damit verbundene Zugangs- oder Nutzungsrechte berücksichtigen. Für die Dokumentation der Objekte hat sich dabei das Referenzmodell CIDOC CRM¹⁰⁵ durchgesetzt, das auch weltweit von Kulturerbeorganisationen und Museen verwendet wird.^{106,107}

¹⁰⁰ Vgl. <http://www.cam.ac.uk/deptdirectory/>; http://thomsonreuters.com/products_services/science/science_products/a-z/arts_humanities_citation_index/#tab2

¹⁰¹ Hull (2006), S. 8

¹⁰² Schäfer; Thänert (2011), S. 3 ff.

¹⁰³ Vom Menschen hergestellte oder benutzte Gegenstände <http://www.archaeological.org/education/glossary#a>

¹⁰⁴ IFA (2008), S. 6

¹⁰⁵ <http://www.cidoc-crm.org/>

¹⁰⁶ Data Documentation and Content Standards <http://ads.ahds.ac.uk/project/userinfo/standards.html>

¹⁰⁷ Dally et al. (2012), S. 169 ff.

Auch die Datentypen und Anwendungen zur Datenverarbeitung innerhalb der Disziplin sind stark heterogen. Es werden unter anderem relationale Datenbanken wie MS Access oder MySQL, Text-Dateien (MS Word, RTF, TXT, PDF), Bilder (JPG, TIF, BMP, GIF), Tabellenkalkulationen, rechnerunterstützte Zeichnungen (CAD) und geographische Daten in GIS-Anwendungen erstellt.¹⁰⁸ Während Feldforschungsprojekten wie Ausgrabungen, Expeditionen, Restaurierungen oder geophysikalischen Prospektionen vor Ort entstandene analoge Dokumentationen werden für die Weiterverarbeitung digitalisiert. Anschließend erfolgt oft eine statistische Auswertung der Daten oder Visualisierung in zwei- und dreidimensionalen Modellen. Daneben werden auch ältere Datenbestände für die historische Forschung retrodigitalisiert.¹⁰⁹

Die präzise Datenerfassung von GIS-Koordinaten stellt ein wichtiges Hilfsmittel in der Archäologie dar. Allerdings dürfen diese oft nur eingeschränkt öffentlich zugänglich gemacht werden, um illegale Grabungen oder Beschädigungen zu vermeiden.¹¹⁰ An das Forschungsdatenmanagement in Repositorien ergeben sich daraus zusätzliche Anforderungen an Zugangskontrollen oder die „Unpräzisierung“ der Daten (z.B. Angabe der Region ohne detaillierten Koordinaten).¹¹¹

Die Bearbeitung der Forschungsfragen in der Archäologie erfolgt unter Verwendung verschiedener Arbeitsmethoden sowohl aus den Geistes- und Sozialwissenschaften, als auch aus den Naturwissenschaften. Diese Interdisziplinärität spiegelt sich in der Breite der Datenformate, die beim Betreiben der Datenrepositorien berücksichtigt werden müssen. Daneben handelt es sich bei einem wesentlichen Teil der Forschungsdaten um physische Artefakte. Die Pflege der digitalen Repräsentation und der physischen Aufbewahrung erfolgt in der Regel gemeinsam durch ein Datenarchiv und ein Museum.

Koordinierung von Forschungsdaten-Aktivitäten

Die Maßnahmen zur Gewährleistung von Interoperabilität und Langzeitverfügbarkeit von Forschungsergebnissen sowie der Technologieeinsatz werden in Deutschland durch das Deutsche Archäologische Institut (DAI) koordiniert.¹¹² Das DAI schreibt unter anderem verbindlich und detailliert vor, welche Datenformate für die Speicherung und die Austauschbarkeit der Forschungsdaten verwendet werden müssen.¹¹³ Für die Integration von Forschungsdaten beim wissenschaftlichen Publizieren ist das sich derzeit noch im Aufbau befindliche Forschungsdatenzentrum IANUS von besonderer Bedeutung. Als nationale Daten- und

¹⁰⁸ Hull (2006), S. 20 ff.

¹⁰⁹ Schäfer; Thänert (2011), S. 2, 5.

¹¹⁰ Meier zu Verl; Horstmann (2011), S. 367

¹¹¹ Hull (2006), S. 14

¹¹² Ebd., S. 6 ff.

¹¹³ DAI IT-Leitfaden, Version 1.0.4. <http://www.dainst.org/de/project/it-leitfaden>

Serviceinfrastruktur soll es zukünftig die Vernetzung und Nachnutzung der archäologischen und altertumswissenschaftlichen Forschungsdaten unterstützen und somit die Bedingungen für *enhanced publications* verbessern.¹¹⁴

Nationale Koordinierungsorganisationen für das Management von Forschungsdaten in der Archäologie sind ebenfalls in Großbritannien, den USA und Australien zu finden. In Großbritannien wurde nach Einstellung der Finanzierung des ehemaligen Arts and Humanities Data Service (AHDS) im Jahr 2008 diese Funktion dem Archaeology Data Service (ADS) übertragen.¹¹⁵ Dieser steht gleichzeitig allen Archäologen Großbritanniens als zentrales Repository offen. Analog dazu ist die US-amerikanische Organisation „Digital Antiquity“ für den Betrieb des zentralen Repository „the Digital Archaeological Record“ (tDAR) und die Langzeiterhaltung der Forschungsdaten zuständig.¹¹⁶ In Australien wird die disziplinübergreifende Koordinierung durch den Australian National Data Service (ANDS) geführt.¹¹⁷ Das disziplinspezifische Datenarchiv „The Australian Historical Archaeological Database“ (AHAD) wird derzeit ebenfalls als ein Bestandteil von tDAR aufgebaut.¹¹⁸

Auf der internationalen Ebene werden Empfehlungen zu bestimmten Verfahren¹¹⁹ und zur guten wissenschaftlichen Praxis¹²⁰ in der Archäologie von verschiedenen Organisationen entwickelt. Die Standards dieser Organisationen sind für deren Mitglieder jeweils verpflichtend und tragen dadurch dazu bei, eine einheitliche Praxis im Umgang mit und Interoperabilität von Forschungsdaten zu etablieren.

3.1.2. Technische Infrastruktur

Untersuchung der Forschungsdaten-Repositoryen

Die Untersuchung der technischen Infrastruktur in der Archäologie erfolgte durch die Auswertung der Funktionalitäten ausgewählter Repositoryen. Um einen repräsentativen Vergleich unter einzelnen Disziplinen zu ermöglichen, wurde in der Vorgehensweise der Masterarbeit eine Mindestzahl von fünf etablierten Repositoryen angestrebt. In den Referenzlisten wurden allerdings für den Fachbereich Archäologie nur jeweils ein bis drei disziplinspezifische Repositoryen aufgelistet. Durch eine weitere Recherche wurde die heterogene Landschaft der oft projektbezogenen Datenbanken und

114 <http://www.dainst.org/forschungsdatenzentrum>

115 <http://archaeologydataservice.ac.uk/about>

116 <http://www.tdar.org/use/stewardship/>

117 <http://www.ands.org.au/>

118 <http://ahad.edu.au/>

119 Vgl. The London Charter zur rechnerunterstützten Visualisierung von Kulturerbe;
<http://www.londoncharter.org/>

120 Vgl. The Institute for Archaeologists <http://www.archaeologists.net/codes/ifa>

Kooperationen mit größeren Datenarchiven deutlich erkennbar. Die technischen Möglichkeiten der Integration von Forschungsdaten wurden schließlich am Beispiel folgender Repositorien untersucht: Archaeology Data Service (ADS)¹²¹, the Digital Archaeological Record (tDAR)¹²², Open Context¹²³, Tree-Ring Database vom World Data Center for Paleoclimatology (Tree-Ring)¹²⁴ und Digital Archaeological Archive of Comparative Slavery (DAACS)¹²⁵. Die Ergebnisse der Auswertung sind in Tabelle 1 zusammengefasst.

Funktionalitäten der Repositorien	ADS	tDAR	Open Context	Tree-Ring	DAACS
Datenformate	PDF, CSV, TXT, XML, SHP, SHX, SBN, PRJ, DBF, JPG, MPG, DXF	JPG, TIF, TXT, XLS, CSV, PDF, OBJ	JPG, PNG, XLS, CSV, PDF	TXT, CRN, RWL, C14, AGE, CHR, SIT, SMP, TIL, WK1, PUB	JPG, XLS, OBJ, SQL
Herunterladen möglich	ohne Anmeldung	Anmeldung erforderlich	ohne Anmeldung	ohne Anmeldung	ohne Anmeldung
Anmeldung erfolgreich	-	ja	-	-	-
PID	DOI	internes ID + DOI	internes ID	internes ID	internes ID
Zitationsvorschlag direkt einbezogen	ja	ja	ja	ja	nein
Version der Forschungsdaten	nein	nein	nein	nein	nein
Methodenbeschreibung	ja	ja	ja	ja	nein
Nutzungsrechte	Terms of Use (akademische Nutzung); CC-Lizenzen	CC-Lizenz (CC-BY)	CC-Lizenz (CC-BY)	Open Access	Open Access
Verwandte Publikationen	ja	ja	ja	ja	nein
Nutzungsstatistik	nein	ja	ja	nein	nein
Deposit-Prozedur	Daten, Lizenz und Dokumentation werden per Email / CD-ROM zugeschickt	Online-Formular	durch Kontaktaufnahme / Einzelfallprüfung (peer review); kostenpflichtig	Online-Formular	via Mitgliedorganisationen

Tabelle 1: Auswertung der Forschungsdaten-Repositorien in der Archäologie

121 <http://archaeologydataservice.ac.uk/archives/>

122 <http://core.tdar.org/>

123 <http://opencontext.org/>

124 <http://www.ncdc.noaa.gov/paleo/treering.html>

125 <http://daacs.org/>

Die untersuchten Repositorien zeichnen sich durch nutzerfreundliche Oberflächen und umfassende Einstiegsmöglichkeiten für Recherchen im Bestand (z.B. nach Titel, Untersuchungszeitraum, vermutetem Datumsbereich, Forschungsprojekt, Untersuchungsmethode, Fundstelle, Kultur, Region, Typen von Artefakten und Materialien) aus. Oft wird die räumliche Eingrenzung (eng. *spatial coverage*) des Fundstückes oder Untersuchungsgebietes auf einer Karte visualisiert. Manche Datensätze im tDAR-Repository wurden nur referenziert und bis zum angegebenen Freigabe-Datum unzugänglich. Somit kann die Einhaltung von Embargofristen und die Sicherung der kommerziellen Interessen beteiligter Parteien gewährleistet werden.¹²⁶ Dabei zeichnen sich die Repositorien durch überwiegend offenen Zugang zu den Forschungsdaten und den Einsatz von CC-Lizenzen aus. Einen hohen Anteil der Bestände stellen Bilder und Fotografien dar – also einmalige Aufnahmen, was keine Versionierung des Datensatzes erklären könnte. Angaben zur letzten Aktualisierung des Eintrages waren hingegen regelmäßig zu finden. Die kodierten Werte werden in der Regel in einem *coding sheet* erklärt (vgl. *codebook* in den Sozialwissenschaften, Kapitel 3.5). Bei DAACS handelte es sich um eine abfragenbasierte Datenbank (eng. *query*), die über keine vorgefertigten Datensätze verfügt, sondern diese nach Benutzervorgaben bildet. Da solche relationale Datenbanken in der Archäologie vergleichsweise oft eingesetzt werden, wurde DAACS zum Zweck der Repräsentativität mit in die Untersuchung aufgenommen.¹²⁷ Bemerkenswert ist noch, dass das vom Alexandria Archive Institute betriebene Open Context-Repository in seiner Beta-Version untersucht wurde und die Funktionalitäten (z.B. PID) noch weiter ausgebaut werden.¹²⁸ Somit soll es der Referenzierbarkeit von Forschungsdaten in der Archäologie noch mehr beitragen. Die facettierte Suche und benutzerfreundliche Navigation in den untersuchten Repositorien entsprechen den in der Literatur oft hervorgehobenen Anforderungen an unkomplizierte und präzise Suchoptionen. Die vergleichsweise eingeschränkte Deposit-Prozedur spiegelt den Umgang mit sensiblen Forschungsobjekten in der Disziplin wider.¹²⁹

Beispiele für Forschungsdatenintegration

Ein Beispiel für die Veröffentlichung von Forschungsdaten mit einer textuellen Dokumentation¹³⁰ stellt das *Journal of Open Archaeology Data* (JOAD)¹³¹ dar. Die Zeitschrift befand sich zum Untersuchungszeitpunkt in der Beta-Version und wird voraussichtlich mit dem ADS kooperieren, das dabei die Archivierung der Forschungsdaten übernehmen wird. Die Datensätze und *data papers* sollen

126 Vgl. Soil Systems, Inc.; Hill, Rebecca (2011) Pueblo Grande DMB Unknown BG demography & paleopathology dataset. (tDAR ID: 367878) <http://dx.doi.org/10.6067/XCV8N87870>

127 Vgl. <http://www.latrobe.edu.au/amc/database.html>

128 <http://opencontext.org/about/people>

129 Vgl. „Open Context only publishes datasets from credentialed researchers working on properly permitted projects” <http://opencontext.org/about/people>; “You will not use content obtained through tDAR in ways that could be reasonably expected to lead, directly or indirectly, to damage to the archaeological record” <http://www.tdar.org/support/policies/terms-of-use/>

130 Vgl. Dallmeier-Tiessen (2011), S. 161

131 <http://openarchaeologydata.metajnl.com/about/>

über eigene DOIs verfügen und gegenseitig verlinkt (eng. *cross-linking*) werden.¹³² Aus der Sicht der *enhanced publications* ist der Fokus auf Nachnutzbarkeit der Forschungsdaten bei dieser Zeitschrift (z.B. durch den Kommentar zum Nachnutzungspotential als Teil des *data papers*) von besonderer Bedeutung. Die Wahrnehmung der präsentierten Forschungsergebnisse wird daneben alternativ zu den etablierten Indikatoren wie Zitationszahlen in bestimmten Datenbanken mit dem *Total Impact*¹³³ gemessen. Als ein von der Fachöffentlichkeit entwickelte Werkzeug könnte dieser künftig einen hohen Potential für Akzeptanz darstellen.

Weitere Versuche zur Integration von Forschungsdaten sind ebenfalls im *Journal of Archaeology in the Low Countries* zu finden. Im dem durch die SURFfoundation geförderten Projekt „Enriched Publications“ wurden unter anderem Möglichkeiten untersucht, interaktive Visualisierungen innerhalb von Artikeln einzubinden.¹³⁴ Allerdings diente das Projekt nur als Pilotverfahren und wurde in die Publikationsprozesse der Zeitschrift nicht übernommen.¹³⁵

3.1.3. Politische Rahmenbedingungen

Im Rahmen der vorliegenden Arbeit sollten nach der Untersuchung der organisationellen und technischen Bedingungen zusätzlich die politischen Vorgaben der jeweiligen Disziplin hinsichtlich der Forschungsdatenintegration untersucht werden. Im Vergleich zu anderen Wissenschaftsdisziplinen wird die Forschung in der Archäologie nicht von ausdrücklich dafür zuständigen Organisationen gefördert, sondern durch ihre Zuordnung den Geistes- und Sozialwissenschaften. So wurden die Anforderungen für Projektförderung nach den Vorgaben im „Research Funding Guide“ des Arts & Humanities Research Council (AHRC)¹³⁶ und in den „Data Management Plans“ des National Endowment for the Humanities (NEH)¹³⁷ untersucht. In Österreich wird die Disziplin unter ihrer Zugehörigkeit zu den Historischen und Kunstwissenschaften gefördert.¹³⁸ Die Auswertung erfolgt hier anhand der allgemeinen „Open Access Policy“ des Fonds zur Förderung der wissenschaftlichen Forschung (FWF).¹³⁹ Die Ergebnisse der Auswertung sind in Tabelle 2 zusammengefasst.

132 Siehe z.B. <http://dx.doi.org/10.5334/4f3bcb3f7f21d>

133 <http://total-impact.org/faq>

134 [https://www.surfgroepen.nl/sites/JALCproject/Project results/Project summary.aspx](https://www.surfgroepen.nl/sites/JALCproject/Project%20results/Project%20summary.aspx)

135 <https://www.surfgroepen.nl/sites/JALCproject/Project%20results/Enhanced%20articles%20in%20JALC.aspx>

136 AHRC (2012): Research Funding Guide. Version 2.0.

<http://www.ahrc.ac.uk/SiteCollectionDocuments/Research-Funding-Guide.pdf>

137 NEH (2012): Data management plans for NEH Office of Digital Humanities Proposals and Awards

http://www.neh.gov/files/grants/data_management_plans_2012.pdf

138 <http://www.fwf.ac.at/de/applications/general/wiss-disz-201002.pdf>

139 FWF (o.J.): Open Access Policy bei FWF-Projekten. http://www.fwf.ac.at/de/public_relations/oai/

Anforderungen der Förderorganisation	FWF	AHRC	NEH
	Österreich	UK	USA
Forschungsdaten müssen öffentlich zugänglich gemacht werden ab	k. A.	k. A.	2012
Zeitraumen	spätestens 2 Jahre nach Projektende	innerhalb von 3 Monaten nach Projektende	zum frühestmöglichen Zeitpunkt
Datenmanagementplan erforderlich	nein	ja	ja
Ausnahmen erlaubt	aus rechtlichen Gründen	bei bestimmten Gründen	für vertrauliche oder geschützte Daten
Zusätzliche Anforderungen	im PubMed gelistete Publikationen müssen im Volltextarchiv UKPMC frei zugänglich gemacht werden ¹⁴⁰	wichtige Datensätze sollten für mindestens 3 Jahre nach dem Förderungsende zugänglich gemacht werden	spezielle Anforderungen an Metadaten und Zugang von GIS-Daten ¹⁴¹
Vorgegebenes Repository	disziplinspezifische oder institutionelle Repositorien	ADS	keine Bevorzugung
Unterstützung angeboten	k. A.	Datenservice, Personal und Guidelines des ADS	Universitäts-einrichtungen oder Fachorganisationen
Sanktionen / Kontrolle	durch Projektberichte	k. A.	durch Projektberichte und Kurzberichte bei den nachstehenden Antragstellungen

Tabelle 2: Anforderungen der Förderorganisationen in der Archäologie

Die untersuchten Förderorganisationen fordern die Forscher grundsätzlich auf, ihre Forschungsergebnisse so früh wie möglich zugänglich zu machen. Dabei werden Forscher ausdrücklich darauf hingewiesen, jegliche Einschränkungen wie beispielsweise eine Embargofrist zu berücksichtigen und diese rechtzeitig der Organisation mitzuteilen. Des Weiteren werden sensible Daten (z.B. personenbezogene Daten oder präzise Ortsangaben der Ausgrabungsstätten) geschützt, indem diese von den Anforderungen ausgeschlossen werden. Außerdem können zusätzliche Anforderungen definiert werden, die sich auf bestimmte Datentypen beziehen. In diesem Fall handelte es bei der Anforderung des NEH um eine regierungsweite Vorschrift. Bei den Feldforschungsprojekten sind gegebenenfalls auch lokale Regelungen zu den Ausgrabungsstätten zu berücksichtigen.

140 Vgl. Publikation aus der vom FWF geförderten mtDNA-Studie
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2829035/> (PubMed),
<http://ukpmc.ac.uk/abstract/MED/20100333> (UKPMC)

141 General Terms and Conditions for Awards, Revised July 2010, Chapter 34:
<http://www.neh.gov/grants/manage/general-terms-and-conditions-awards-awards-issued-may-2009-or-later#geographic>

3.2. Klimaforschung

3.2.1. Organisation und Arbeitsweise der Disziplin

Forschungsdaten in der Disziplin

Die Klimaforschung stellt zusammen mit den Nachbarbereichen aus den Marinewissenschaften, satellitengestützter Erdbeobachtung und der Untersuchung der Erdoberfläche einen Teil der Forschungsdisziplinen des Erdsystems dar.¹⁴² Der Fachbereich befasst sich größtenteils mit der Dokumentation der Zustände des Klimas unter Verwendung verschiedener Messgeräte (u.a. zur Temperatur, Luftdruck und -feuchtigkeit, Windstärke, Saison-Erscheinungen wie Pollenflug) und der Modellierung der komplexen Systeme für Prognosen zum Klimawandel. Durch den großen Einfluss ihrer Ergebnisse auf weitreichende Entscheidungen in Medizin, Politik, Tourismus und anderen Bereichen hat die Klimaforschung eine hohe gesellschaftliche Bedeutung.¹⁴³

Die Genauigkeit und Vertrauenswürdigkeit der Daten spielt eine besondere Rolle in der Klimaforschung. Ihre Ergebnisse stehen derzeit vor allem im Zentrum des politischen Interesses. Um diese Anforderungen zu erfüllen, werden die meteorologischen Messstationen weltweit verteilt und erzeugen Daten teilweise im Minuten-Takt. Für verlässliche Aussagen zum Klimawandel werden mehrere Modellierungen mit verschiedenen Modifikationen benötigt, die anschließend statistisch ausgewertet werden müssen. Bedingt durch diese Arbeitsweise werden in der Disziplin umfangreiche und komplexe Daten generiert.¹⁴⁴

Außerdem müssen die klimarelevanten Daten über Jahrzehnte hinweg aufbewahrt werden, um eine historische Analyse der Klimaveränderung zu ermöglichen.¹⁴⁵ Dies bringt zugleich Anforderungen an die Langzeitarchivierung und Standardisierung der verwendeten Formate mit sich. So haben sich unter den Klimawissenschaftlern Formate wie NetCDF¹⁴⁶, GRIB¹⁴⁷ und HDF¹⁴⁸ für detaillierte Repräsentation disziplinspezifischer Daten etabliert. Zusätzlich sind alle Messdaten mit den Koordinaten zur geographischen Länge (eng. *longitude*) und Breite (eng. *latitude*) am Erhebungsort zu verbinden. Diese Angaben basieren oft auf den Standards ISO 19115 und ISO 19139.¹⁴⁹

¹⁴² Vgl. <http://www.dkrz.de/daten-en/wdcc/wdc-cluster-earth-system-research>

¹⁴³ Fritzsche (2012), S. 195, 211

¹⁴⁴ Fritzsche / Ebd., S. 196

¹⁴⁵ Fritzsche / Ebd., S. 207

¹⁴⁶ <http://badc.nerc.ac.uk/help/formats/netcdf/>

¹⁴⁷ <http://www.grib.us/>

¹⁴⁸ <http://eosweb.larc.nasa.gov/HBDOCS/hdf.html>

¹⁴⁹ Vgl. Fritzsche (2012), S. 207

Von einem bewussten Umgang mit Forschungsdaten in der Disziplin zeugt ebenfalls das vom International Council for Science (ICSU) koordinierte Netzwerk der Weltdatenzentren (World Data Centres, WDCs).¹⁵⁰ Das überwiegend den Geowissenschaften und den dazugehörigen Unterdisziplinen gewidmete System soll die Bereitstellung und Langzeitverfügbarkeit der hochwertigen Daten sichern.¹⁵¹ Nicht zuletzt hat sich auch das DataCite-System zur Zitation der Forschungsdaten aus einem DFG-geförderten Projekt mit den Weltdatenzentren Climate, MARE, RSAT und dem GFZ Potsdam entwickelt.¹⁵²

Die Datenerhebung in der Disziplin erfolgt unter Verwendung diverser Geräte. Die Beobachtungsdaten stammen oft aus den Messstationen staatlicher Einrichtungen,¹⁵³ das großräumige Monitoring der Erde erfolgt durch Satelliten, für die Messungen innerhalb der Atmosphäre werden Ballonaufstiege durchgeführt, in der Paläoklimaforschung werden Eiskernbohrungen eingesetzt und Jahresringe der Bäume analysiert.¹⁵⁴ Schließlich werden für großvolumige Simulationen des Klimas Hochleistungsrechner eingesetzt.¹⁵⁵ Während die großen Volumina von erzeugten Daten eher homogene Formate aufweisen, zeichnen sich die kleineren Volumina oft durch die Heterogenität der Formate aus.¹⁵⁶

Für die Auswertung von Daten wird Software für statistische Berechnungen und Visualisierungen wie MATLAB oder R benutzt. Die raumbezogenen Daten werden dabei mittels ArcGIS-Anwendungen¹⁵⁷ erfasst.¹⁵⁸ Großvolumigen Datensätze können vor dem Herunterladen aus Datenarchiven beispielsweise mittels einer Jblob-Software¹⁵⁹ in kleinere Einzeldateien aufgeteilt werden.¹⁶⁰

Die Datentypen in der Klimaforschung lassen sich in einmalige observationelle Daten und Daten aus den Simulationen mit den Modellen aufgliedern. Aufgrund des oft sehr großen Umfangs der Daten werden diese nicht von einzelnen Organisationen, sondern in den für ein bestimmtes Forschungsbereich spezialisierten Repositorien archiviert. Daher wird die in der vorliegenden Arbeit empfohlene Vorgehensweise für *enhanced publications* – also Verlinkung von Zeitschriftenartikeln mit Forschungsdaten in fachspezifischen Repositorien – aus vielerlei Gründen als eine optimale Lösung in der Disziplin angesehen.

150 <http://icsu-wds.org/our-members/membership-categories/regular-members>

151 <http://icsu-wds.org/organization/intro-to-wds>

152 Fritzsche (2012), S. 208

153 Vgl. Deutscher Wetterdienst am Bundesministerium für Verkehr, Bau und Stadtentwicklung
<http://www.dwd.de/>

154 Vgl. <http://www.ncdc.noaa.gov/paleo/paleo.html>

155 Fritzsche (2012), S. 202-203

156 Vgl. Lautenschlager (2011), S. 151; Fritzsche (2012), S. 203

157 <http://www.esri-germany.de/products/arcgis/index.html>

158 Vgl. Hamann (2012), S. 260

159 <http://cera-www.dkrz.de/CERA/jblob/>

160 Vgl. DKRZ empfiehlt die Nutzung von Jblob ab der Dateigröße von 4 GB

Koordinierung von Forschungsdaten-Aktivitäten

Die Kollaboration und gute wissenschaftliche Praxis in der Disziplin wird auf der internationalen Ebene von der Weltorganisation für Meteorologie (World Meteorological Organisation, WMO) koordiniert.¹⁶¹ Als ein Organ der UN legt sie unter anderem standardisierte Prozeduren fest, die für alle Mitglieder der Organisation verbindlich gelten.¹⁶² Daneben werden umfangreiche Richtlinien zu den technischen und organisationellen Aspekten im Umgang mit Forschungsdaten¹⁶³ und Empfehlungen zu den Datenformaten angeboten.¹⁶⁴

Für die Zusammenarbeit und den Wissensaustausch in der Forschung des Klimawandels stellt das Intergovernmental Panel on Climate Change (IPCC)¹⁶⁵ eine zentrale Institution dar. Dabei werden technische Richtlinien zum Datenmanagement¹⁶⁶ und Szenarien zu den Arbeitsabläufen¹⁶⁷ entwickelt, die für die Einhaltung der guten Praxis insbesondere im Bereich der Klimamodellierung beitragen. Die einheitliche Verwendung von Metadaten wird in Anlehnung an die NetCDF Climate and Forecast (CF) Metadata Convention bestrebt.¹⁶⁸ Das Abkommen definiert zusätzliche Komponenten für das etablierte NetCDF-Format, um den Austausch und die Vergleichbarkeit der Daten aus den verschiedenen Quellen zu erleichtern. Eine solche Anreicherung der Metadaten bindet die Angaben zum Kontext der Datenerhebung mit ein und stellt somit eine wichtige Grundlage für die spätere Nachnutzung der Daten dar.¹⁶⁹

Alternativ zu den fachspezifischen Organisationen spielen auch nationale Ansätze wie das Australian Research Collaboration Service (ARCS)¹⁷⁰ eine gewichtige Rolle.¹⁷¹ Es werden u.a. Datendienste für die Speicherung und Zugänglichmachung von Forschungsdaten und kollaborative Arbeitsumgebungen angeboten. Durch den benutzerfreundlich aufgebauten und kostenlosen Service wird das Ziel verfolgt, die gute Praxis im Forschungsdatenmanagement zu etablieren und das *data sharing* in allen Wissenschaftsdisziplinen in Australien zu unterstützen.¹⁷²

161 www.wmo.int/

162 Vgl. WMO (1988), S. 5

163 http://www.wmo.int/pages/prog/www/WIS/manuals_guides_techregulations_en.html

164 http://www.wmo.int/pages/prog/sat/formatsandstandards_en.php

165 <http://www.ipcc.ch/>

166 <http://www.ipcc-data.org/guidelines/index.html>

167 http://www.ipcc-data.org/ddc_envdata.html

168 <http://cf-pcmdi.llnl.gov/>

169 Vgl. http://badc.nerc.ac.uk/help/formats/netcdf/index_cf.html

170 <http://www.arcs.org.au/>

171 Vgl. Hense; Quadtr (2011)

172 <http://www.arcs.org.au/index.php/arcs-data-fabric>

3.2.2. Technische Infrastruktur

Untersuchung der Forschungsdaten-Repositoryen

Die Untersuchung der technischen Infrastruktur in der Klimaforschung erfolgte durch die Auswertung der Funktionalitäten ausgewählten Repositoryen. Als repräsentativ für den Fachbereich wurden fünf Messdaten-Repositoryen ausgewählt, die in mindestens zwei Referenzlisten aufgelistet worden waren: British Atmospheric Data Centre (BADC)¹⁷³, Canada's National Climate Data and Information Archive¹⁷⁴, National Snow and Ice Data Center (NSIDC)¹⁷⁵, PANGAEA - Data Publisher for Earth & Environmental Science¹⁷⁶ und World Data Center for Climate (WDCC)¹⁷⁷. Die Ergebnisse der Auswertung sind in Tabelle 3 zusammengefasst.

Funktionalitäten der Repositoryen	BADC	Canada's National Climate Archive	NSIDC	PANGAEA	WDCC
Datenformate	HDF, NetCDF, NASA-Ame, INO, ICC, HAL, HAS, AWS, DIF, JPG, GIF, XLS, CSV, DAT, TXT, ASC, PS, PDF, BPD, OPJ, DOC	CSV, XML, TXT	ASC, AUX, TIF, GIF, TXT, GPD, DAT, BIN, ASC, KML, DIF	TAB, TXT	NetCDF, GRB, XML, JPG
Herunterladen möglich	teilweise Anmeldung erforderlich	ohne Anmeldung	Anmeldung erwünscht aber nicht erforderlich	ohne Anmeldung	Anmeldung erforderlich
Anmeldung erfolgreich	nein	-	-	-	ja
PID	internes ID	internes ID	internes ID	DOI	DOI, URN
Zitationsvorschlag direkt einbezogen	ja	nein	ja	ja	ja
Version	nein	nein	nein	nein	nein
Methodenbeschreibung	ja	nein	ja	ja	ja
Nutzungsrechte	für akademische Nutzung	Open Data	Open Data	CC-Lizenz (CC-BY)	CC-Lizenz (CC-BY-NC-SA)
Verwandte Publikationen	ja	nein	ja	ja	ja
Nutzungsstatistik	nein	nein	nein	nein	nein

¹⁷³ <http://badc.nerc.ac.uk/>

¹⁷⁴ <http://climate.weatheroffice.gc.ca/>

¹⁷⁵ <http://nsidc.org/data/data-search.html>

¹⁷⁶ <http://www.pangaea.de/>

¹⁷⁷ <http://cera-www.dkrz.de/WDCC/ui/Index.jsp>

Deposit-Prozedur	Online-Formular; nur für registrierte Nutzer aus Partner- organisationen	Daten werden durch nationale Messstation en ermittelt	Online- Formular	Online- Formular	durch Kontakt- aufnahme / Einzelfall- prüfung; kostenpflichtig
------------------	---	--	---------------------	---------------------	--

Tabelle 3: Auswertung der Forschungsdaten-Repositoryn in der Klimaforschung

Die untersuchten Repositoryn zeichneten sich vor allem durch eine individuelle Themenabgrenzung und der dafür geeigneten Datenformate aus. Das kanadische Datenarchiv fungiert dabei als eine Art der nationalen Anstalt zur Beobachtung des Klimas. Als solche verfolgt es hauptsächlich das Ziel, die Messdaten zu archivieren und verfügbar zu machen und bietet somit keine weiteren Funktionalitäten an.

Alle anderen Repositoryn hingegen zeugen von der hohen Bedeutung von Forschungsdaten im Publikationsprozess. Alle vier beziehen den Zitationsvorschlag für den Datensatz in der Beschreibung direkt ein und verweisen zugleich auf verwandte Publikationen, die auf Basis dieser Daten entstanden sind. Die Angabe der bibliographischen Daten eines (geplanten) Artikels wird insbesondere bei der Einreichung von Forschungsdaten in PANGAEA gefördert.

Allerdings versehen nur zwei der untersuchten Repositoryn die Forschungsdaten mit persistenten Identifikatoren. Die Angaben zur Version der Daten werden in keinem Repository hinterlegt. Alternativ wird stattdessen das Datum der Erhebung oder der letzten Aktualisierung verwendet. Dies könnte dadurch erklärt werden, dass die Mess- und Beobachtungsdaten meist nur einmal an einem bestimmten Ort (Messstation) und zum bestimmten Zeitpunkt erhoben werden und über keine weitere Aufarbeitung verfügen. Gleichzeitig werden dadurch oft sehr große Datenmengen erzeugt, so dass z.B. die Beantragung eines DOI für jeden Datensatz nicht immer realisierbar sein kann.

Im Gegensatz zur überwiegend freien Zugänglichkeit und Nachnutzung der Daten unterscheiden sich die Repositoryn durch ihre uneinheitlichen Bedingungen für die Aufnahme der Daten in den Bestand. Während NSIDC und PANGAEA für alle Wissenschaftler weltweit offen stehen, werden vom BADC und WDCC nur Daten aus bestimmten Organisationen beziehungsweise zu bestimmten Fragestellungen akzeptiert. Dies verdeutlicht wiederum die Spezialisierung von Forschungsdaten-Repositoryn in der Disziplin.

Schließlich werden in allen Repositoryn mehrere Möglichkeiten für die Recherche im Datenbestand angeboten. Als eine etablierte Praxis kann dabei der Suchverlauf nach den geographischen Angaben zum Erhebungsort, dem Erhebungszeitraum oder nach dem Projekttitle genannt werden. Zusätzlich

werden die geographischen Koordinaten der Daten oft auf einer Weltkarte visualisiert. Dadurch ist es auch möglich, über die Markierung des Gebietes auf der Karte oder direkte Angabe der Koordinaten nach den dort erhobenen Forschungsdaten zu suchen.

Beispiele für Forschungsdatenintegration

Im Bereich der Klimaforschung lassen sich bereits einige Initiativen zur Anerkennung der Datenpublikation als wissenschaftlicher Leistung beobachten. Die Anforderungen für die Umsetzung eines prototypischen *Journal for Meteorological Data* wurden in dem von JISC geförderten Projekt „Overlay Journal Infrastructure for Meteorological Sciences (OJIMS)“ untersucht.¹⁷⁸ Folgend den Empfehlungen der Projektgruppe¹⁷⁹ hat die Royal Meteorological Society (RMets) den Start eines Geoscience Data Journal (GDJ) angekündigt.¹⁸⁰ Die publizierten *data papers* in diesem Journal sollen kurze Hintergrundinformationen zu den beschriebenen Datensätzen liefern. Die Datensätze selbst werden dabei in anerkannten Datenzentren archiviert und mittels eines DOI von der Zeitschrift verlinkt.¹⁸¹

Ein weiteres Beispiel einer Veröffentlichung von Forschungsdaten mit einer textuellen nicht-interpretativen Dokumentation¹⁸² stellt das Earth System Science Data (ESSD) dar.¹⁸³ Die zu publizierenden Datensätze werden einem qualitativen *peer review*-Verfahren unterzogen und müssen in vertrauenswürdigen Datenzentren abgelegt werden.¹⁸⁴ Mit der Unterstützung von persistenten Identifikatoren, barrierefreier Zugänglichkeit durch das Open Access-Modell und klaren Nutzungsbedingungen durch CC-Lizenzen tragen beide Journale zur offenen Publikationskultur und zum *data sharing* in der Disziplin bei.

Auch unter den bereits existierenden wissenschaftlichen Zeitschriften sind Beispiele von Forschungsdatenintegration zu finden. So leistet PANGAEA als eines der von ESSD empfohlenen Datenrepositorien¹⁸⁵ gleichzeitig einen Beitrag für *enhanced publications*, indem Artikel des wissenschaftlichen Verlegers Elsevier und deren zugrundeliegende Datensätze beidseitig verlinkt werden.¹⁸⁶ Diese seit 2010 bestehende Zusammenarbeit zur Fortentwicklung des wissenschaftlichen Publizierens dient sogar als ein Musteransatz für Datenrepositorien in anderen

178 <http://www.jisc.ac.uk/whatwedo/programmes/reppres/sue/ojims.aspx>

179 Callaghan et al. (2009), S. 21

180 Pressemitteilung vom 16. Juli 2012, <http://www.rmets.org/new-geoscience-data-journal>

181 <http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%292049-6060/homepage/ProductInformation.html>

182 Vgl. Dallmeier-Tiessen (2011), S. 162

183 <http://www.earth-system-science-data.net/>

184 http://www.earth-system-science-data.net/general_information/about_this_journal.html

185 http://www.earth-system-science-data.net/general_information/repository_criteria.html

186 http://www.elsevier.com/wps/find/authored_newsitem.cws_home/companynews05_01616

Wissenschaftsdisziplinen.¹⁸⁷ Allerdings sind die mit Forschungsdaten angereicherte Zeitschriftenartikel aufgrund der kommerziellen Interessen des Verlegers vorwiegend nur kostenpflichtig zugänglich¹⁸⁸ und bieten somit keinen vollständigen Zugang zu den Forschungsergebnissen.

Eine Maßnahme zur Standardisierung der Publikationsabläufe wird daneben im Rahmen des durch die DFG geförderten Projektes „Publikation Umweltdaten“ entwickelt.¹⁸⁹ Mittels einer Software wird die automatische Überprüfung nach Abweichungen innerhalb der Forschungsdaten durchgeführt und ein Bericht im XML-Format erstellt.¹⁹⁰ Basierend auf diesem Ablauf werden die meteorologischen Daten und deren Metadaten für die Überführung in das WDCC kontrolliert und die Publikation mit DataCite vorbereitet.¹⁹¹ Obwohl die Verantwortung für die wissenschaftliche Qualität der Daten dem Autor überlassen bleibt, lässt sich durch den Einsatz automatischer Fehlererkennung die technische Qualität der archivierten Daten steigern.¹⁹²

3.2.3. Politische Rahmenbedingungen

Die Förderung der Klimaforschung erfolgt durch ihre Zuordnung zur Großgruppe der Geo- oder Umweltwissenschaften. Die politischen Rahmenbedingungen der Disziplin wurden demzufolge nach den Vorgaben drei Förderorganisationen untersucht: die „Open Access Policy“ der Niederländischen Organisation für Wissenschaftliche Forschung (NWO)¹⁹³, die „Data Policy“ des National Environment Research Council (NERC)¹⁹⁴ und die „Data Sharing Policy“ der National Oceanic and Atmospheric Administration (NOAA)¹⁹⁵. Die Fördermittel der Abteilung für Geo- und Lebenswissenschaften (ALW)¹⁹⁶ der NWO werden dabei in Anlehnung an allgemeine NWO-Regelungen vergeben.¹⁹⁷ Die Ergebnisse der Auswertung sind in Tabelle 4 zusammengefasst.

Anforderungen der Förderorganisation	NWO-ALW	NERC	NOAA
	Niederlande	UK	USA
Forschungsdaten müssen öffentlich zugänglich gemacht werden ab	Mai 2011	Januar 2011	Oktober 2011

187 http://wiki.datadryad.org/Elsevier_Integration

188 Siehe z.B. <http://dx.doi.org/10.1016/j.rse.2010.11.018>

189 <http://umwelt.wikidora.com/wikidora/>

190 Hense, Quad (2011), [ohne Seitennummer], <http://dx.doi.org/10.1045/january2011-hense>

191 <http://cera-www.dkrz.de/atarrabi2/home.seam>

192 Fritzsche (2012), S. 204

193 http://www.nwo.nl/nwohome.nsf/pages/NWOP_89BBXM_Eng

194 <http://www.nerc.ac.uk/research/sites/data/policy2011.asp>

195 <https://www.nosc.noaa.gov/EDMC/PD.DSP.php>

196 http://www.nwo.nl/nwohome.nsf/pages/ACPP_4WMJGH_Eng

197 http://www.nwo.nl/nwohome.nsf/pages/SPES_5VEDDR_Eng

Zeitraumen	zum frühestmöglichen Zeitpunkt	innerhalb von 2 Jahren nach der Datenerhebung	innerhalb von 2 Jahren nach der Datenerhebung
Datenmanagementplan erforderlich	nein	ab Juni 2012	ja
Ausnahmen erlaubt	kommerzielle und vertrauliche Daten	zum Schutz der Rechte Dritter	aus datenschutzrechtlichen Gründen
Zusätzliche Anforderungen	Leistungsschutzrecht bzw. einfaches Nutzungsrecht für NWO	Data Value Checklist für Bestimmung von LZA	Veröffentlichung des DMP in Verbindung mit Forschungsdaten
Vorgegebenes Repository	keine Bevorzugung	Bestimmte Repositorien	keine Bevorzugung
Unterstützung	durch NWO und Fachorganisationen	durch das Repository	Mustervorlage für DMP und eigenes Repository geplant
Sanktionen / Kontrolle	durch Projektberichte	Einstellen der Zahlungen oder Ausschließen aus der Förderung künftiger Projekte	Erfüllung des DMP wird bei den nachstehenden Antragstellungen berücksichtigt

Tabelle 4: Anforderungen der Förderorganisationen in der Klimaforschung

Derzeit wird von der NWO in Kooperation mit DANS eine Open Access Richtlinie für Forschungsdaten entwickelt.¹⁹⁸ Zum Untersuchungszeitpunkt befand sich diese allerdings noch im Aufbau. Die Position dieser Förderorganisation war bereits aus der allgemeinen Open Access-Richtlinie ersichtlich, dürfte aber in der nächsten Zeit weitere Änderungen oder konkretere Anforderungen einbringen.

Die Richtlinien von NERC und NOAA wiesen dabei vergleichsweise strenge Vorgaben auf. Kennzeichnend dafür sind die zusätzlichen Anforderungen beider Organisationen und die Kontrollmaßnahmen für die Erfüllung der Anforderungen. Die historische Bedeutung der Forschungsdaten für langfristige Rückschlüsse auf Klimaveränderungen ist ebenfalls aus den Vorgaben des NERC erkennbar. So müssen in den Förderprojekten entstandene Forschungsdaten nach dem *Data Value Checklist* bewertet werden, um potentiell hochwertige Daten für die Langzeitarchivierung zu identifizieren.

¹⁹⁸ NWO (2010)

Des Weiteren wird von den Förderorganisationen zwischen den Forschungsdaten und den daraus entstandenen Mehrwertdienstleistungen (eng. *value added services*) unterschieden.¹⁹⁹ Von den Anforderungen für Förderung in der Klimaforschung werden somit oft als kommerzielle Dienstleistungen umgesetzte Informationsprodukte ausgeschlossen. Dies veranschaulicht nochmals die Rolle der Forschungsdaten in der Wertschöpfungskette der Informationsversorgung für Wissenschaft und Gesellschaft.

3.3. *Kristallographie*

3.3.1. Organisation und Arbeitsweise der Disziplin

Forschungsdaten in der Disziplin

Die Entstehung der modernen Kristallographie wird oft mit dem Einsatz der Röntgenstrahlung (eng. *X-ray*) für die Bestimmung von dreidimensionalen Strukturen auf der atomaren und molekularen Ebene verbunden.²⁰⁰ Auf diese Weise gewonnene Erkenntnisse werden von benachbarten Disziplinen der Chemie, Physik, Biologie oder den Geowissenschaften angewendet. Die Methoden der Kristallographie werden sogar im Bereich des Kulturerbes eingesetzt.²⁰¹ Aufgrund ihres direkten Bezuges auf die Bestimmung der Materialeigenschaften liefert die Kristallographie zudem eine Grundlage für Produktentwicklungen in der Industrie.^{202,203}

Die Datentypen in dieser Disziplin umfassen größtenteils Zeichnungen, Tabellen, Bilder und Textdokumente um. Eine besondere Art der Daten stellen daneben die spektroskopischen Daten zu den Energiezuständen in Kristallen (eng. *spectrum, spectra*) (z.B. im JDX-Format)²⁰⁴ und deren graphische Abbildungen dar.²⁰⁵ Die räumliche Anordnung der Atome wird dabei hauptsächlich mittels Röntgenbeugungsmethode (eng. *X-ray diffraction*) visualisiert.²⁰⁶ Die Erschließung von Kristallstrukturen in den Forschungsdaten-Repositorien beinhaltet dabei spezielle Metadaten zur chemischen Formel, Eigenschaften der Struktur (z.B. Atomkoordinaten, Dichte, Molekülmasse) oder Rahmenbedingungen während der Durchführung von Experimenten (Temperatur, Druck, Historie der Probenahme).²⁰⁷ Die Interoperabilität der Daten wird durch den im Fachbereich etablierten CIF-

199 Vgl. NERC Data policy statement <http://www.nerc.ac.uk/research/sites/data/policy2011.asp#n1>

200 <http://www.iucr.org/iycr>

201 Vgl. <http://www.iucr.org/iucr/commissions/art-and-cultural-heritage> ;
http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503478

202 <http://www.iucr.org/iycr>

203 Polydoratos (2006), S. 8, 27

204 http://badc.nerc.ac.uk/help/formats/jcamp_dx/

205 Polydoratos (2006), S. 14 ff.

206 Ebd., S. 27

207 Grazulis et al. (2012), S. D421

Standard gewährleistet.²⁰⁸ Des Weiteren beziehen sich die Datensätze in der Regel nicht auf einzelne Strukturen der Atome oder Moleküle, sondern auf bestimmte Verbindungen derselben (eng. *compounds*).²⁰⁹ Für die Nutzung in gedruckten und elektronischen Ressourcen werden komplexe chemische Substanzen mittels International Chemical Identifier (InChI)²¹⁰ und Simplified molecular-input line-entry system (SMILES)²¹¹ auf vereinfachte Weise repräsentiert.²¹²

Für die Einbindung von kristallographischen Forschungsdaten beim wissenschaftlichen Publizieren sind detaillierte Metadaten zu den Eigenschaften der untersuchten Struktur notwendig. Außerdem werden in der Disziplin oft gemeinsame Projekte mit den Partnern aus der Industrie (z.B. Chemie-Konzerne) durchgeführt. Im Rahmen solcher Projekte entstandene Daten stellen möglicherweise die Vorarbeiten für die Patentanmeldung oder Produktentwicklung dar. Zum Schutz kommerzieller Interessen werden solche Forschungsdaten oft nur für die Projektgruppe zugänglich gemacht. Auch diese Einschränkungen müssen bei der Arbeit mit Forschungsdaten der Kristallographie berücksichtigt werden.

Koordinierung von Forschungsdaten-Aktivitäten

Der effiziente Umgang mit Forschungsdaten und gute wissenschaftliche Praxis in der Disziplin wird auf der internationalen Ebene von der International Union of Crystallography (IUCr)²¹³ vorangetrieben. Die IUCr setzt sich dabei insbesondere für die Integrität und Zuverlässigkeit der Forschungsergebnisse. In einem öffentlichen Brief an die Herausgeber wissenschaftlicher Zeitschriften forderte sie beispielsweise die Validierung der Publikationen zugrundeliegenden Datensätze auf.²¹⁴ Die Nutzung des eigen entwickelten checkCIF-Services soll damit zur Vorbeugung wissenschaftlichen Fehlverhaltens beitragen, indem die beschriebenen Strukturen nach deren Übereinstimmung und Vollständigkeit automatisch überprüft werden.²¹⁵ Bei konkreten Maßnahmen zum Forschungsdatenmanagement verweist die IUCr allerdings auf das Committee on Data for Science and Technology (CODATA) des International Council for Science (ICSU).²¹⁶ Das Komitee bemüht sich insbesondere um effizienten Umgang mit Forschungsdaten in naturwissenschaftlichen Disziplinen. In einer speziellen Arbeitsgruppe von CODATA werden unter anderem Empfehlungen zur Zitierbarkeit

208 Crystallographic Information Framework (CIF); Vgl. <http://www.iucr.org/resources/cif>

209 Vgl. <http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=1183>

210 <http://www.iupac.org/home/publications/e-resources/inchi.html>

211 <http://en.wikipedia.org/wiki/SMILES>

212 <http://www.chemspider.com/Chemical-Structure.13860434.html>

213 <http://www.iucr.org/>

214 IUCr (2011): Publication standards for crystal structures,
<http://www.iucr.org/home/leading-article/2011/2011-06-02>

215 <http://checkcif.iucr.org/>

216 <http://www.iucr.org/resources/data>

der Forschungsdaten entwickelt. Dabei verweist die Arbeitsgruppe auf die Initiativen von DataCite und den Einsatz von persistenten Identifikatoren als geeignete Vorgehensweisen hin.²¹⁷

Auf nationaler Ebene wird in Großbritannien ein Service für Bestimmung der Kristallstrukturen aufgebaut. Der an der University of Southampton angesiedelte National X-ray Crystallography Service (NCS) soll ausgewählte Kristalle mittels eigener leistungsfähiger Geräte wie Diffraktometer und Synchrotron untersuchen, um die dabei gewonnenen Daten anschließend im eCrystals Repository zu archivieren.²¹⁸ Somit soll es auch anderen Forschern ermöglicht werden, die Strukturdaten referenzierbar und öffentlich zugänglich zu machen.²¹⁹ Die Erfassung von Daten kann dabei vollständig durch den NCS übernommen werden, indem sogar publikationsreife Abbildungen und Tabellen als ein Teil der Serviceleistung geliefert werden können.²²⁰ Unter dem Begriff „service crystallography“ werden analoge Dienstleistungen auch von anderen wissenschaftlichen Einrichtungen angeboten.²²¹ Im Kontext solcher nationalen Initiativen werden die Fragen zur Zitierung von Forschungsdaten und der Einhaltung guter wissenschaftlicher Praxis möglicherweise zukünftig noch mehr an Bedeutung gewinnen.

Die Beantwortung der Forschungsfragen in der Kristallographie setzt kollaborative Strukturen verschiedener Organisationen voraus. Dies ist vor allem bedingt durch die Abhängigkeit von leistungsfähigen Großgeräten, die nur in ausgewählten Einrichtungen zur Verfügung stehen und von verschiedenen Forschern und Institutionen gemeinsam genutzt werden. Aufgrund der Lieferung von Basisinformationen für weitergehende Produktentwicklungen in der Chemie und der Industrie werden hohe Anforderungen an die Zuverlässigkeit der Forschungsergebnisse gestellt. Die Erfüllung dieser Anforderungen wird mittels verschiedener Standardisierungsverfahren durch Koordinationsorganisationen unterstützt.

3.3.2. Technische Infrastruktur

Untersuchung der Forschungsdaten-Repositoryen

Die Forschungsdaten-Infrastruktur in der Kristallographie wurde durch die Auswertung der Funktionalitäten ausgewählter Repositoryen untersucht. Als repräsentativ für den Fachbereich wurden fünf Kristallstrukturen-Repositoryen ausgewählt, die in mindestens zwei Referenzlisten geführt

217 <http://www.codata.org/taskgroups/TGdatacitation/>

218 National X-ray Crystallography Service (NCS), S. 1 ff.
<http://www.epsrc.ac.uk/SiteCollectionDocuments/other/XRAY.pdf>

219 <http://ecrystals.chem.soton.ac.uk/>

220 <http://www.ncs.ac.uk/services/data.html>

221 Vgl. http://www.staff.tugraz.at/christoph.marschner/service_crystallography.html

werden. Dazu gehören die Cambridge Structural Database (CSD)²²², ChemSpider²²³, die Crystallography Open Database (COD)²²⁴, eCrystals²²⁵ und PubChem²²⁶. Die CSD stellt dabei das Hauptprodukt des Cambridge Crystallographic Data Centre (CCDC) dar, das von der University of Cambridge betrieben wird.²²⁷ Allerdings ist für die Installation der CSD die Erwerbung einer kostenpflichtigen Lizenz notwendig. Die Funktionalitäten dieser Datenbank wurden daher unter Verwendung einer Demo-Version des WebCSD-Portals²²⁸ untersucht. Die Ergebnisse der Auswertung sind in Tabelle 5 zusammengefasst.

Funktionalitäten der Repositorien	CSD	ChemSpider	COD	eCrystals	PubChem
Datenformate	CIF, FCF, HKL, SDF, MOL2, TXT	CIF, JDX, CDX, MOL, SDF, PNG, JPG, GIF, SKC	CIF, HKL, TXT	CIF, CML, FCF, RES, PRP, HKL, JPG, GIF, HTM, DOC, InChI, INS, MOL, LST	ASN, XML, SDF, InChI, CSV, Pc3D, JPG, PNG
Herunterladen möglich	lizenzpflichtig	ohne Anmeldung	ohne Anmeldung	ohne Anmeldung	ohne Anmeldung
Anmeldung erfolgreich	demo login	-	-	-	-
PID	internes ID	internes ID	internes ID	internes ID + DOI	internes ID
Zitationsvorschlag direkt einbezogen	ja	ja	nein	ja	nein
Version	nein	nein	nein	nein	nein
Methodenbeschreibung	ja	ja	ja	ja	ja
Nutzungsrechte	für akademische Nutzung	Open Data	Open Access	Open Data	für akademische Nutzung
Verwandte Publikationen	ja	ja	ja	nein	ja
Nutzungsstatistik	nein	nein	nein	nein	nein
Deposit-Prozedur	Online-Formular	Online-Formular	Online-Formular	Online-Formular	Online-Formular

Tabelle 5: Auswertung der Forschungsdaten-Repositorien in der Kristallographie

Unter den untersuchten Repositorien zeichnete sich das von der Royal Society of Chemistry²²⁹ betriebene ChemSpider durch umfangreiche Recherchemöglichkeiten aus. Die Daten zu chemischen Verbindungen in dem Repository werden von mehr als 400 Ressourcen aggregiert und über das Portal

222 <http://www.ccdc.cam.ac.uk/products/csd/>

223 <http://www.chemspider.com/>

224 <http://www.crystallography.net/>

225 <http://ecrystals.chem.soton.ac.uk/>

226 <http://pubchem.ncbi.nlm.nih.gov/>

227 http://www.ccdc.cam.ac.uk/about_ccdc/

228 http://www.ccdc.cam.ac.uk/products/csd_system/webcsd/

229 <http://www.rsc.org/>

als eine zentrale *one-stop-shop*-Schnittstelle mit verwandten Publikationen in Fachzeitschriften, Patenten und Einträgen in Fachdatenbanken einschließlich deren externen Identifikatoren direkt verlinkt. Daneben stellt es umfassende Einstiegsmöglichkeiten zur Verfügung. So können beispielsweise die Kristallstrukturen über das Zeichnen der Teilstruktur der Moleküle gesucht werden.²³⁰ Ähnliche Funktionalitäten bietet ebenfalls das vom National Center for Biotechnology Information²³¹ betriebene PubChem.

Allgemein war der freie Zugang zu den Forschungsdaten im Sinne von Open Access deutlich erkennbar. Eingeschränkte Nutzungsbedingungen der CSD stellten damit eine Ausnahme in der Disziplin dar. Grundsätzlich dominiert eine Kultur des offenen *data sharing*, was beispielsweise durch eine vom COD initiierte „Petition for Open Data in Crystallography“ gegen kostenpflichtige Datenbanken belegt wird. Die Initiative vertritt das Prinzip, das die Information zur räumlichen Anordnung der Atome als ein Fakt nicht urheberrechtlich geschützt werden kann und rief in der Vergangenheit unter anderem Wissenschaftler dazu auf, die CIF-Dateien an das COD zu spenden. Diese Petition soll zum Anstieg des COD-Bestands erheblich beigetragen haben.²³²

Beispiele für Forschungsdatenintegration

Als Antwort auf den Aufruf von IUCr zur Validierung und Sicherung von den Artikeln zugrundeliegenden Forschungsdaten bieten sowohl freie als auch kostenpflichtige Repositorien Autoren an, CIF-Dateien mit den Strukturfaktoren zu archivieren. Teilweise wird dies auch von Seiten der Journale gefordert.²³³ Dadurch sind in der Disziplin oft Publikationen anzutreffen, in denen zugrundeliegende Strukturdaten und weitere unterstützende Materialien öffentlich zugänglich gemacht werden, ohne den Zugriff auf die eigentliche interpretative Publikation zu gewähren.²³⁴ Solche Artikel können somit als eine Art von *data papers* in der Disziplin angesehen werden.

Des Weiteren lassen sich verschiedene Versuche zur Integration von Forschungsdaten beobachten. So werden beispielsweise die Strukturdaten zu dem Kristall in manchen Zeitschriften sowohl als Primärdaten im CIF-Format, als auch als *supporting information* in zusammengefasster Form im PDF-Format angeboten. Die Daten können dabei oft über den Verleger der Zeitschrift uneingeschränkt heruntergeladen werden.²³⁵ Die unterschiedlichen Vorgehensweisen beim Publizieren von Forschungsdaten zeugen damit eher von der Erprobung der technischen Möglichkeiten als von der Existenz fest etablierter allgemeiner Praktiken in diesem Bereich.²³⁶

230 <http://www.chemspider.com/About.aspx>

231 <http://www.ncbi.nlm.nih.gov/>

232 <http://www.crystallography.net/petition/>

233 <http://beta-www.ccdc.cam.ac.uk/Community/Depositastructure/Pages/StructureDepositionInformation.aspx>

234 Siehe z.B. <http://dx.doi.org/10.1107/S1600536801016427>

235 Siehe z.B. <http://dx.doi.org/10.4125/pd0025th>

236 TIB et al. (2010), S. 38 ff.

Generell sind die wissenschaftlichen Publikationen in der Kristallographie sehr datenintensiv. Der Schwerpunkt liegt meist auf der Beschreibung der Eigenschaften der Struktur und den Bedingungen während des Experimentes. Die Interpretation der Forschungsergebnisse stellt dabei in der Regel nur einen kleinen Teil des Artikels dar. Die meisten Publikationen in der Disziplin können somit als *enhanced publications* bezeichnet werden, sobald beide Bestandteile – der beschreibende Artikel und die Primärdaten – uneingeschränkt zugänglich gemacht werden.

3.3.3. Politische Rahmenbedingungen

Die Forschung in der Kristallographie wird durch ihre Zuordnung dem Fachbereich der Chemie und dem Großbereich der Lebens- oder Naturwissenschaften gefördert. Da eine Untersuchung der Förderorganisationen nach den übergeordneten Bereichen keine präzise Informationen für die Förderbedingungen in der Kristallographie liefern kann, wurden nur diejenigen Organisationen berücksichtigt, welche Forschungsprojekte oder Repositorien in der Kristallographie in der Vergangenheit bereits gefördert haben. Betrachtet wurden somit das „Policy Framework on Research Data“²³⁷ des Engineering and Physical Sciences Research Council (EPSRC) in Großbritannien, „Advice to PIs on Data Management Plans“²³⁸ des NSF Directorate of Mathematical and Physical Sciences, Division of Chemistry in den USA, die „NHMRC revised policy on the dissemination of research findings“²³⁹ des National Health and Medical Research Council sowie der Australian Code for the „Responsible Conduct of Research“²⁴⁰ in Australien. Die Ergebnisse der Auswertung sind in Tabelle 6 zusammengefasst.

Anforderungen der Förderorganisation	NHMRC	EPSRC	NSF-CHE
	Australien	UK	USA
Forschungsdaten müssen öffentlich zugänglich gemacht werden ab	k. A.	Mai 2011	Januar 2011
Zeitraumen	innerhalb von 6 Monaten nach der Erhebung	zum frühestmöglichen Zeitpunkt; Metadaten innerhalb von 12 Monate nach der Erhebung	zum frühestmöglichen Zeitpunkt
Datenmanagementplan erforderlich	nein	nein	ja

237 <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/default.aspx>

238 <http://www.nsf.gov/bfa/dias/policy/dmpdocs/che.pdf>

239 <http://www.nhmrc.gov.au/grants/policy/dissemination-research-findings>

240 NHMRC (2007)

Ausnahmen erlaubt	in begründeten Fällen	in begründeten Fällen	vertrauliche oder geschützte Daten
Zusätzliche Anforderungen	Sicherung der LZA für mindestens 5 Jahre durch die Forschungsinstitution	Sicherung der LZA für mindestens 10 Jahre; Nutzung von PID	5 Fragen für DMP
Vorgegebenes Repository	institutionelle oder fachspezifische Repositories	institutionelle oder fachspezifische Repositories	keine Bevorzugung
Unterstützung	Institution der Wissenschaftler	Hinweis auf JISC und DCC	k. A.
Sanktionen / Kontrolle	durch Projektberichte	Einzelfallprüfung	k. A.

Tabelle 6: Anforderungen der Förderorganisationen in der Kristallographie

Nach der überarbeiteten und seit Juli 2012 gültigen Richtlinie des NHMRC müssen alle aus den geförderten Projekten entstandenen wissenschaftlichen Artikel in einem institutionellen Repository im Sinne von Open Access frei zugänglich gemacht werden. Die Organisation sieht vor, die Embargofrist der Journale von maximal 12 Monaten zu berücksichtigen, so dass die ersten Publikationen ab dem 1. Juli 2013 frei zugänglich gemacht werden müssen. Der offene Zugang zu Forschungsdaten innerhalb von 6 Monaten nach deren Erhebung wird dabei ausdrücklich empfohlen, stellt aber keine verbindliche Vorgabe dar. Allerdings weist die Organisation darauf hin, dass die Forscher dazu verpflichtet sind, die Forschungsergebnisse unterstützenden Primärdaten zu sichern. Die Verantwortung dafür kommt insbesondere den Forschungseinrichtungen zu, indem jede wissenschaftliche Einrichtung unter anderem über formale Richtlinien zur Aufbewahrung von Forschungsdaten verfügen muss.²⁴¹

Analog dazu sieht auch EPSRC vor, alle durch Förderprojekte entstandenen wissenschaftlichen Artikel öffentlich zugänglich zu machen.²⁴² Die Forschungsdaten sollen dabei über ihre Metadaten referenzierbar gemacht werden und mit persistenten Identifikatoren wie DOI versehen werden.²⁴³ Der offene Zugang zu den Forschungsdaten selbst wird somit nicht vorausgesetzt. Allerdings werden die Forschungsinstitutionen dazu verpflichtet, nach eigenen Richtlinien die Langzeitarchivierung der Forschungsdaten zu sichern und bei begründeten Anfragen zur Verfügung zu stellen.

²⁴¹ Ebd., S. 2.1 ff.

²⁴² <http://www.epsrc.ac.uk/about/infoaccess/Pages/roaccess.aspx>

²⁴³ <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/expectations.aspx>

Im Allgemeinen vermitteln die Vorgaben der untersuchten Förderorganisationen ein positives Klima für *enhanced publications*. Es wird grundsätzlich erwartet, die Forschungsergebnisse öffentlich zugänglich zu machen. Dies betrifft sowohl wissenschaftliche Zeitschriftenartikel, als auch Forschungsdaten selbst. Außerdem wird empfohlen, bei Publikationen auf die zugrundeliegenden Forschungsdaten und Zugangsmöglichkeiten zu diesen hinzuweisen. Schließlich werden die etablierten Praktiken in den Wissenschaftsdisziplinen anerkannt, indem die Wahl der geeigneten Vorgehensweise dem Wissenschaftler selbst überlassen bleibt.

3.4. Ökologie

3.4.1. Organisation und Arbeitsweise der Disziplin

Forschungsdaten in der Disziplin

Die Wissenschaftsdisziplin der Ökologie widmet sich der Erforschung von Wechselwirkungen zwischen der biologischen Diversität und deren Funktion im Ökosystem. In ihren Fragestellungen werden insbesondere die Auswirkungen der Handlungen vom Menschen und der damit verbundene Verlust von Arten im Ökosystem untersucht.²⁴⁴ Die Forschungsdaten in der Disziplin entstehen primär über Messungen und die Analyse von Böden und Bodenorganismen, Pflanzen, Tieren und biochemischen Stoffkreisläufen. Außerdem werden Landnutzungsbefragungen und Waldinventuren durchgeführt, Fotografien, Karten und Modelle erstellt sowie Fernerkundungs- und GIS-Daten erhoben. Während der Feldarbeiten (sog. Explorationen) gesammelte Forschungsdaten wie Bodenproben werden anschließend im Labor ausgewertet.²⁴⁵

Eine Besonderheit der Forschungsdaten und somit auch der Metadatenschemata in diesem Fachbereich stellt die Einordnung der Arten (sog. Spezies) in eine biologische Klassifikation dar. Die Arten werden nach der hierarchischen Systematik in deren Gattung, Familie, Ordnung, Klasse, Stamm, Reich und Lebewesen eingeordnet.²⁴⁶ Eine solche systematische Einheit bildet demgemäß ein Taxon ab.²⁴⁷ Für die räumliche und zeitliche Zuordnung von Spezies und deren Lebensräumen (sog. Habitaten) sowie eine einheitliche Terminologie in den Metadaten haben sich Darwin Core (DwC)²⁴⁸ und Access to Biological Collections Data (ABCD)²⁴⁹ als Standards etabliert.²⁵⁰ Weiterhin werden die

244 Vgl. <http://kn.b.ecoinformatics.org/biodiversity/index.jsp>

245 Vgl. Nieschulze; König-Ries (2011), S. 216 ff.

246 Siehe z.B. <http://www.morphbank.net/Browse/ByTaxon/?tsn=39838>

247 Vgl. International Code of Botanical Nomenclature (2006): Division II, Chapter I, Article 3: <http://ibot.sav.sk/icbn/main.htm>

248 <http://www.tdwg.org/activities/darwincore/>

249 <http://www.tdwg.org/activities/abcd/>

fachspezifischen Metadaten auf Basis der Ecological Metadata Language (EML)²⁵¹ und in dem für die Naturwissenschaften entwickelten NeXus und NeXML-Format²⁵² definiert.²⁵³

Die Forschungsdaten im Fachbereich beziehen sich zum großen Teil auf die textuelle oder multimediale Repräsentation der untersuchten Organismen.²⁵⁴ Dies können beispielsweise numerische Daten aus Laborexperimenten, Audio- und Videodateien von Tierbeobachtungen, Fotografien oder rechnerunterstützte Simulationen sein.²⁵⁵ Die Text-Dateien liegen oft im ASCII-Format vor und werden mit den gängigen Text-Editoren bearbeitet.²⁵⁶ Die Bilder und Zeichnungen werden ebenso in den konventionellen Formaten wie JPG, TIF, BMP, GIF gespeichert.²⁵⁷ Die nicht-firmeneigenen (eng. *non-proprietary*) Formate²⁵⁸ werden daneben auch beim Publizieren von Forschungsdaten bevorzugt oder sogar gefordert. Somit sprechen sich die Datenarchive für die maximale Zugänglichkeit der publizierten Forschungsergebnisse aus.²⁵⁹

Koordinierung von Forschungsdaten-Aktivitäten

Der freie Zugang zu den Forschungsdaten und deren Vernetzung im Bereich der Biodiversität ist eines der strategischen Ziele des Global Earth Observation System of Systems (GEOSS) für das Jahr 2015.²⁶⁰ Durch die Conservation Commons-Initiative haben ebenfalls mehrere Organisationen den drei Prinzipien zum offenen und rechtmäßigen Zugang zu Forschungsdaten zugestimmt.²⁶¹

Internationale Zusammenarbeit und Interoperabilität der Biodiversitätsdaten werden vom „Biodiversity Information Standards“ (TDWG, ehem. Taxonomic Databases Working Group) vorangetrieben.²⁶² Als ein an der International Union of Biological Sciences angegliederter Verein ist dieser insbesondere für die Entwicklung der einheitlichen Standards zuständig. Dazu gehören u.a. bereits etablierte Standards wie Darwin Core und ABCD.²⁶³

Die Publikation von Forschungsdaten in Übereinstimmung mit den Standards wird durch Software wie das Integrated Publishing Toolkit des Global Biodiversity Information Facility (GBIF) unterstützt. So

250 Vgl. Nieschulze; König-Ries (2011), S. 222

251 <http://knb.ecoinformatics.org/software/eml/>

252 <http://www.nexusformat.org/>

253 Vgl. <http://treebase.org/treebase-web/search/study/summary.html?id=2182>

254 Vgl. Chavan; Penev (2011), S. 2

255 Vgl. Dallmeier-Tiessen (2011), S. 157

256 Vgl. Nieschulze; König-Ries (2011), S. 219

257 Vgl. Biegon (2006), S. 16

258 Von der Nutzung kommerzieller Software unabhängige Formate; vgl. Jantz (2003)

259 Vgl. http://esapubs.org/archive/instruct_d.htm

260 http://www.earthobservations.org/geoss_ta_da_tar.shtml

261 <http://conservationcommons.net/about-2/>

262 <http://www.tdwg.org/about-tdwg/>

263 <http://www.tdwg.org/standards/>

werden die Biodiversitätsdaten vor der Veröffentlichung über das GBIF-Netzwerk nach der Konformität mit DarwinCore und EML überprüft.²⁶⁴ Auf nationaler Ebene²⁶⁵ und in benachbarten Disziplinen wie der Ornithologie²⁶⁶ sind ebenfalls Initiativen zur Entwicklung und Umsetzung von *best practices* im Umgang mit digitalen Forschungsdaten zu beobachten. So werden beispielsweise Datenmodelle und Grundprinzipien entwickelt, die den Datenaustausch im jeweiligen Sachgebiet unterstützen sollen.

3.4.2. Technische Infrastruktur

Untersuchung der Forschungsdaten-Repositorien

Die Untersuchung der technischen Infrastruktur in der Ökologie erfolgte durch die Auswertung der Funktionalitäten ausgewählter Repositorien. Als repräsentativ für den Fachbereich wurden fünf Repositorien ausgewählt, die Daten für Biodiversitätsforschung enthalten und in mindestens zwei Referenzlisten aufgelistet worden waren: Dryad²⁶⁷, Knowledge Network for Biocomplexity (KNB)²⁶⁸, Morphbank²⁶⁹, Paleobiology Database²⁷⁰ und TreeBASE²⁷¹. Die Ergebnisse der Auswertung sind in Tabelle 7 zusammengefasst.

Funktionalitäten der Repositorien	Dryad ²⁷²	KNB	Morphbank	Paleobiology Database	TreeBASE
Datenformate	XLS, TXT, XML, STR, FASTA (FAS), PDF, DOC, PHY, BIO, U, FNA, NEX, NXS, SDD, CON	TXT, DAT, XLS, CSV, KML, XML	JPG, TIF	XLS, CSV	TXT, NEX, NeXML
Herunterladen möglich	ohne Anmeldung	teilweise Anmeldung erforderlich	ohne Anmeldung	ohne Anmeldung	ohne Anmeldung
Anmeldung erfolgreich	-	ja	-	-	-
PID	DOI	internes ID	internes ID	internes ID	internes ID + URI

²⁶⁴ <http://www.gbif.org/informatics/infrastructure/publishing/#c889>

²⁶⁵ Vgl. National Biodiversity Network <http://www.nbn.org.uk/The-NBN/What-is-the-NBN/NBN-Concept.aspx>

²⁶⁶ Vgl. Martín, Ballard (2010)

²⁶⁷ <http://datadryad.org/>

²⁶⁸ <http://knb.ecoinformatics.org/data.jsp>

²⁶⁹ <http://www.morphbank.net/>

²⁷⁰ <http://paleodb.org/>

²⁷¹ <http://treebase.org/>

²⁷² Über Dryad wird auch ein Teil von KNB und TreeBASE-Beständen referenziert

Zitationsvorschlag direkt einbezogen	ja	ja	nein	nein	nein
Version	nein	nein	nein	nein	nein
Methoden- beschreibung	ja	ja	ja	ja	ja
Nutzungsrechte	CC-Lizenzen (CC0); Open Data	Nutzungsvertrag; Kopien der Publikationen müssen dem Archiv zugesandt werden	CC-Lizenzen (CC-BY-NC- SA)	frei zugänglich	frei zugänglich
Verwandte Publikationen	ja	nein	nein	ja	ja
Nutzungsstatistik	ja	nein	nein	nein	nein
Deposit-Prozedur	Online- Formular	Online-Formular	Online- Formular	Online- Formular; nur für registrierte Mitglieder	Online- Formular

Tabelle 7: Auswertung der Forschungsdaten-Repositories in der Ökologie

Innerhalb der untersuchten Repositories fällt das Dryad durch seine Spezialisierung auf die Artikeln zugrundeliegende Forschungsdaten auf. Deshalb verfügt dieses Repository über wesentliche für *enhanced publications* Funktionalitäten wie beispielsweise persistente Identifizierung und freie Zugänglichkeit der Forschungsdaten sowie über „motivationssteigernde“ Elemente wie der direkte Zitationsvorschlag, Nutzungsstatistik und Hinweis auf die auf den Daten basierenden wissenschaftlichen Publikationen. Manche Datensätze im Dryad-Repository waren bedingt durch Embargofrist bis zum bestimmten Zeitpunkt unzugänglich. Dies stellt dabei einen wichtigen Mechanismus für die Kontrolle über die Daten und die Übereinstimmung mit den Policies der Journale dar.

Bei den Forschungsdaten in der Paleobiology Database und TreeBASE handelte es sich größtenteils um Taxa aus älteren Zeitschriftenartikeln. Diese sind für Forscher von historischer Bedeutung und stellen eine Art „Retro-Digitalisierung“ von Forschungsdaten in dem Fachbereich dar. Daneben haben alle Repositorien explizit das Interesse an den aus den Forschungsdaten entstehenden Publikationen verkündet.²⁷³ Des Weiteren können die Daten von den Archiven überwiegend frei heruntergeladen werden und lassen damit eine offene *data sharing*-Kultur erkennen. Außerdem sind Möglichkeiten zur Darstellung verschiedener Rollen bei der Datenerhebung deutlich erkennbar. So können beispielsweise die Kontaktdaten des Datenmanagers einer zuständigen Einrichtung oft identifiziert werden. Dies könnte als Beleg für den bewussten Umgang mit Forschungsdaten und gut organisierte Strukturen dienen.

273 Vgl. <http://www.morphbank.net/About/Manual/submit.php>

Beispiele für Forschungsdatenintegration

Durch die Joint Data Archiving Policy des Dryad-Repository und seine Partnerschaft mit den wissenschaftlichen Zeitschriften²⁷⁴ verfügt der Fachbereich der Ökologie und Biowissenschaften bereits über mehrere *enhanced publications*, deren zugrundeliegende Forschungsdaten frei nachnutzbar und nachprüfbar sind. Die Autoren werden dabei oft verpflichtet, die Forschungsdaten bei der Einreichung des Manuskripts für Gutachter zugänglich zu machen. Im Falle der Annahme zur Publikation müssen die Forschungsdaten über geeignete Repositorien anschließend auch der Fachöffentlichkeit frei zugänglich gemacht werden.²⁷⁵

Ebenfalls existieren eigenständige nicht-interpretative Publikationsformen von Forschungsdaten. Von der Ecological Society of America (ESA) werden beispielsweise „Ecological Archives“ herausgegeben, in denen die Forschungsdaten aus den anderen Zeitschriften des Vereins als sog. *data papers* erscheinen.²⁷⁶ Im Rahmen des EU FP-7 Projektes ViBRANT²⁷⁷ wird aktuell das „Biodiversity Data Journal“ gegründet.²⁷⁸ Durch die Kooperation mit den Forschungsdatenrepositorien soll es Autoren angeboten werden, die Datensätze als *data papers* ohne Größeneinschränkungen zu veröffentlichen.²⁷⁹ Der Verleger „Pensoft Publishers“ setzt sich dabei für den Einsatz von Semantic Web-Architekturen und die Vernetzung von Forschungsdaten aktiv ein.²⁸⁰

3.4.3. Politische Rahmenbedingungen

Die Bedingungen für die Förderung in der Ökologie werden im Folgenden hinsichtlich ihrer Fachzugehörigkeit zu einer übergeordneten Disziplin der Biologie untersucht. Betrachtet wurden somit die „Data sharing policy“ des Biotechnology and Biological Sciences Research Council (BBSRC)²⁸¹ und „Information about the Data Management Plan“ des Directorate for Biological Sciences von NSF.²⁸² In Deutschland wird derartige Forschung dem Großbereich der Lebenswissenschaften zugeordnet²⁸³ und stellt damit den größten Förderungsbereich der DFG dar.²⁸⁴ Die Position der DFG

274 <http://datadryad.org/jdap>

275 Vgl. <http://www.nature.com/authors/policies/availability.html>

276 <http://esapubs.org/archive/>

277 <http://vibrant.eu/>

278 Vishvas; Penev (2011), S. 9

279 <http://www.pensoft.net/journals/bdj>

280 <http://www.pensoft.net/page.php?P=23>

281 <http://www.bbsrc.ac.uk/organisation/policies/position/policy/data-sharing-policy.aspx>

282 <http://www.nsf.gov/bio/pubs/BIODMP061511.pdf>

283 Vgl. DFG Förderatlas 2012, Tabelle 2-3: DFG-Systematik der Fachkollegien, Fachgebiete und Wissenschaftsbereiche
http://www.dfg.de/download/pdf/dfg_im_profil/evaluation_statistik/foerderatlas/tabellen/dfg-foerderatlas2012_tab_2-3.xls

lässt sich dennoch nur aus drei verschiedenen Dokumenten zusammenstellen: Vorschläge zur Sicherung guter wissenschaftlicher Praxis: Empfehlungen der Kommission Selbstkontrolle in der Wissenschaft²⁸⁵, Leitfaden für die Antragstellung der Projektanträge²⁸⁶ und Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten²⁸⁷ des Unterausschusses für Informationsmanagement. Das letzte Dokument soll dabei als ein disziplinübergreifender Rahmen für zukünftige Fördermaßnahmen der DFG und die Entwicklung der weiteren disziplinspezifischen Regelungen dienen.²⁸⁸ Die Ergebnisse der Auswertung sind in Tabelle 8 zusammengefasst.

Anforderungen der Förderorganisation	DFG	BBSRC	NSF-BIO
	Deutschland	UK	USA
Forschungsdaten müssen öffentlich zugänglich gemacht werden ab	1998	April 2007	Januar 2011
Zeitraumen	nach Abschluss der Forschungen oder nach wenigen Monaten	innerhalb von 3 Jahren nach der Datenerhebung	zum frühestmöglichen Zeitpunkt
Datenmanagementplan erforderlich	ja; Konzept für die Überführung der nachnutzbaren Forschungsdaten in vorhandene Datenbanken oder Repositorien soll dem Antrag beigelegt werden	ja	ja
Ausnahmen erlaubt	vertrauliche oder kommerzielle Daten	vertrauliche oder nicht-nachnutzbare Daten	vertrauliche oder kommerzielle Daten
Zusätzliche Anforderungen	den Veröffentlichungen zugrundeliegende Daten sollen in der Forschungsinstitution für 10 Jahre aufbewahrt werden; Berücksichtigung der internationalen und fachspezifischen Standards	Aufbewahrung in der Forschungsinstitution 10 Jahre nach dem Projektende; Berücksichtigung der etablierten Standards und Formate	5 Fragen für DMP
Vorgegebenes Repository	institutionelle oder fachspezifische Repositorien	Empfehlungen gegeben aber nicht verpflichtend	keine Bevorzugung
Unterstützung	Universitätseinrichtungen oder Fachorganisationen	Beratung durch BBSRC	k. A.

284 34,9% der DFG-Bewilligungen für 2008 bis 2010; DFG Förderatlas 2012, Tabelle 2-14: Förderbasierte Wissenschaftsindikatoren im Überblick: Anteile nach Wissenschaftsbereichen

http://www.dfg.de/download/pdf/dfg_im_profil/evaluation_statistik/foerderatlas/tabellen/dfg-foerderatlas2012_tab_2-14.xls

285 http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf

286 DFG-Vordruck 54.01 – 6/12

http://www.dfg.de/foerderung/antragstellung/hinweise_vorbereitung_antraege/index.html#micro5217702

287 http://www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf

288 Ebd., S. 1 ff.

Sanktionen / Kontrolle	k. A.	durch Projektberichte; Erfüllung des DMP kann bei nachstehenden Antragstellungen berücksichtigt werden	durch Projektberichte und Kurzberichte bei den nachstehenden Antragstellungen
------------------------	-------	--	--

Tabelle 8: Anforderungen der Förderorganisationen in der Ökologie

Die Empfehlungen der DFG zur Sicherung guter wissenschaftlicher Praxis sind bereits 1998 erschienen. Allerdings handelt es sich dabei um eine durch wissenschaftliches Fehlverhalten veranlasste Denkschrift und keine verbindlichen Vorgaben. Der Leitfaden für Projektantragstellung und die späteren Empfehlungen sprechen ebenfalls nur für die Nachnutzung geeigneter Forschungsdaten an. Die Entscheidung über den Wert der Daten bleibt dabei dem Wissenschaftler selbst überlassen.²⁸⁹ Im Gegensatz dazu dürfen durch die Förderung von BBSRC und NSF entstandene Forschungsdaten nur in begründeten Ausnahmefällen nicht veröffentlicht werden. Die Nachnutzbarkeit der Daten wird allerdings ebenso von dem BBSRC berücksichtigt.

Im Allgemeinen lassen sich die Vorgaben der Förderorganisationen als vergleichsweise flexibel bezeichnen. Dies könnte möglicherweise durch sensible Fragestellungen der Disziplin erklärt werden, bei denen eine sofortige Veröffentlichung von Forschungsdaten nicht immer wünschenswert ist. Der Schwerpunkt im Umgang mit Forschungsdaten der Disziplin liegt nicht in deren Einbindung in wissenschaftliche Publikationsprozesse, sondern vielmehr in der Verlässlichkeit der Forschungsergebnisse. Die Verantwortung für die Langzeitarchivierung ist für die Forschungseinrichtung des Wissenschaftlers somit deutlich erkennbar. Als eine grundlegende Bedingung für die Nachprüfbarkeit der Forschungsergebnisse wird dadurch die Verpflichtung der Hochschulen und Forschungsorganisationen zur Einhaltung der guten wissenschaftlichen Praxis hervorgehoben.

3.5. Sozialwissenschaften

3.5.1. Organisation und Arbeitsweise der Disziplin

Forschungsdaten in der Disziplin

Die Forschung in den Sozialwissenschaften beruht stark auf quantitativen Arbeitsmethoden und empirischen Daten. Die wichtigsten Datenquellen stellen amtliche Statistiken²⁹⁰ und Ergebnisse aus

²⁸⁹ Ebd.

²⁹⁰ Vgl. Amtliche Nachrichten der Bundesagentur für Arbeit (ANBA)
<http://www.pub.arbeitsamt.de/hst/services/anba/index.shtml>

Befragungen dar.²⁹¹ Als qualitative Forschungsprimärdaten können daneben Interviews, Tagebücher, anthropologische Feldnotizen²⁹² oder Paradata²⁹³, Bilder, Fotografien, Audio- und Videoaufnahmen, Berichte und Transkriptionen verwendet werden.²⁹⁴ Des Weiteren können die Forschungsdaten aus Mikro- und Makrodaten globaler Datenbanken wie beispielsweise von Weltbank, UNESCO oder OECD²⁹⁵ generiert werden.²⁹⁶ In der vorliegenden Arbeit wird sich hauptsächlich auf empirisch arbeitende Sozialwissenschaften bezogen.

Aufgrund der Fragestellungen der Disziplin wird es oft mit sensiblen, vertraulichen oder personenbezogenen Daten gearbeitet. Dies bringt besondere Anforderungen an Diskretion und den Schutz der Privatsphäre der Studienteilnehmer mit sich. In wissenschaftlichen Arbeitsabläufen widerspiegelt es sich durch Anonymisierung bestimmter Daten.²⁹⁷ Außerdem wird die Re-Identifikation der Probanden bei Abschluss des Nutzungsvertrages von den meisten Datenarchiven ausdrücklich unterlassen.²⁹⁸

Eine zentrale Datenerhebungsmethode für große nationale und internationale Studien stellt Längsschnittanalyse (eng. *longitudinal analysis*) dar. Während solcher Studien werden bestimmte Daten zu unterschiedlichen Zeitpunkten erhoben (in sog. Wellen) und die Ergebnisse miteinander verglichen.²⁹⁹ Da solche Forschungsdaten nur im Zusammenhang mit anderen Wellen einen interpretativen Vergleich zu den Veränderungen im Laufe der Zeit ermöglichen, müssen diese auch in Datenarchiven miteinander verknüpft werden können. Koordinierende Organisationen haben zu diesem Zwecke spezielle Verfahren entwickelt.³⁰⁰

Eine weitere Besonderheit im wissenschaftlichen Arbeitsprozess ist die Gebundenheit der Disziplin an die Studien-Ebene. Die Forschungsergebnisse werden oft im Kontext einer größeren Untersuchung präsentiert statt einzeln publiziert zu werden. Auch im Bereich der Forschungsdaten bedeutet es eine höhere Granularität, indem beispielsweise DOIs nicht den einzelnen Tabellen zugeteilt werden, sondern der ganzen Studie, bestehend aus primären oder verarbeiteten Datensätzen, Fragebögen, Methodenbeschreibungen und weiteren dazugehörigen Dokumenten.³⁰¹ Eine Ausnahme in dieser

291 Huschka et al. (2011), S. 35 ff.

292 Informationen zu dem Kontext, den Akteuren und dem allgemeinen Verlauf der beobachteten Situation, oft informell und subjektiv; vgl. Friebertshäuser (2010), S. 313

293 Informationen zu den Bedingungen während der Datenerhebung; vgl. Rasmussen (2008), S. 91

294 ICPSR (2012), S. 11, 27 ff.

295 Vgl. <http://www.nsd.uib.no/macrodataloguide/source.html>

296 http://www.nsd.uib.no/macrodataloguide/macro_data.html

297 Huschka et al. (2011), S. 35

298 Vgl. Blankovertrag §2, Punkt 6 <http://psychdata.zpid.de/index.php?main=take&sub=vertrag>

299 Vgl. European Social Survey (ESS) <http://www.europeansocialsurvey.org/> [Zugriff am 26.07.2012]

300 Vgl. Hansen et al. (2011), S. 1

301 Vgl. European Commission (2012): Eurobarometer 75.1 (2011). TNS OPINION & SOCIAL, Brussels. GESIS Data Archive, Cologne. Study ZA5479, Data file Version 5.0.0, <http://dx.doi.org/10.4232/1.11413>

Praxis stellt die OECD iLibrary dar, in der einzelne Teile der Publikationen oft mit eigenen DOIs registriert werden.³⁰²

Die inhaltliche Dokumentation der sozialwissenschaftlichen Daten erfolgt oft in Anlehnung an das Metadatenschema der Data Documentation Initiative (DDI).³⁰³ Die Metadaten umfassen dabei üblicherweise die Beschreibung von Population, Grundgesamtheit, Geographischem Raum, Referenzzeitraum, Auswahlverfahren, Erhebungsverfahren, Frequenz und zeitlicher Dimension der Studie.³⁰⁴ Der einzelne Datensatz besteht ferner aus Einheiten (z.B. Individuum, Organisation, Haushalt, Ereignis)³⁰⁵ und Variablen (jedes Element der analysierten Einheit wie Alter, Einkommen oder Verbraucherpreisindex).³⁰⁶ Die Anzahl der Einheiten und Variablen wird in der Regel bei der Erfassung des Datensatzes mit angegeben. Beide Elemente stellen somit eines der wichtigsten Arbeitsobjekte in den empirischen Sozialwissenschaften dar.

Die Bearbeitung der Daten erfolgt meistens mittels Software für statistische Analysen wie SPSS, STATA, SAS, R oder auch speziell für rechnerunterstützte Befragungen (eng. *computer-assisted interviewing*, CATI/CAPI) entwickelte Software wie Blaise und CASES.³⁰⁷ Die Ergebnisse aus den Umfragen müssen für die Weiterverarbeitung vorab kodiert werden. Zu den standardisierten Verfahrensweisen gehören z.B. die Kodierung der geographischen Einheiten nach ISO 3166-2/3³⁰⁸ oder der Todesursachen nach dem International Classification of Diseases (ICD).³⁰⁹ Damit eine korrekte Interpretation und die Nachnutzung solcher Forschungsdaten möglich ist, müssen sonstige benutzerdefinierte Kodierungen in einem Code-Buch (eng. *codebook*) festgelegt und zusätzlich der Methodenbeschreibung hinzugefügt werden.³¹⁰

Koordinierung von Forschungsdaten-Aktivitäten

Eine zentrale Institution für Koordinierung der Forschungsdaten-Infrastruktur in Deutschland ist der Rat für Sozial- und Wirtschaftsdaten (RatSWD). Die Akkreditierung von thematisch organisierten sozialwissenschaftlichen Forschungsdatenzentren und Datenservicezentren stellt eine der wichtigsten Aufgaben des RatSWD dar.³¹¹ Im Unterschied zu anderen Wissenschaftsdisziplinen sind diese

302 Vgl. OECD (2009), Table A5.1b. Enjoyment of science, in: Education at a Glance, OECD Publishing.
http://dx.doi.org/10.1787/eag-2009-tablea5_1b-en

303 <http://www.ddialliance.org/>

304 Hausstein et al. (2011), S. 16 ff.

305 http://www.ddialliance.org/Specification/DDI-CV/AnalysisUnit_1.0.0.html

306 <http://www.icpsr.umich.edu/icpsrweb/ICPSR/help/glossary.jsp>

307 ICPSR (2012), S. 19, 40 ff.

308 http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=39718

309 <http://www.who.int/classifications/icd/>

310 ICPSR (2012), S. 23

311 <http://ratswd.de/rat/aufgaben.php>

Datenzentren nicht unbedingt der Organisationseinheit gleichgestellt.³¹² So betreibt beispielsweise GESIS – Leibniz-Institut für Sozialwissenschaften - das größte Institut für Sozialwissenschaften in Deutschland, gleichzeitig vier solche Datenzentren.³¹³

Von besonderem Interesse für die Referenzierbarkeit und Zitierbarkeit der Forschungsdaten ist das von GESIS durchgeführte und durch die DFG geförderte Projekt zum Aufbau einer DOI-Registrierungsagentur „da|ra“. Nach Abschluss der Pilot- und der Ausbauphase soll es den deutschen Datenarchiven, Organisationen und Einzelproduzenten von Forschungsdaten ermöglicht werden, ihre Daten mit einem DOI zu registrieren.³¹⁴ Die technische Interoperabilität der Dienstleistung soll dabei durch Kooperation mit DataCite und in Anlehnung an den DDI-Standard gewährleistet werden.³¹⁵

Andere nationale Service für Registrierung und persistente Identifizierung sozialwissenschaftlicher Daten bieten der Economic and Social Data Service (ESDS)³¹⁶ des UK Data Archive in Großbritannien und die Data Archiving and Networked Services (DANS)³¹⁷ in den Niederlanden. Für *enhanced publications* sind beide Organisationen durch ihre Beteiligung an speziellen Projekten von besonderem Interesse. So spielte das UK Data Archive eine wesentliche Rolle bei der Entwicklung einer technischen Spezifikation für die Verlinkung von Forschungsdaten und Literatur im StORe-Projekt.³¹⁸ Daneben beteiligte sich DANS an der Vorführung von Prototypen für *enhanced publications* im DRIVER-II-Projekt.³¹⁹ Die Realisierung von *enhanced publications* wird sogar als eines der Ziele der DANS-Strategie zum 2015 vorgesehen.³²⁰

Internationale Initiativen zur Koordinierung von Forschungsdaten haben sich in den Sozialwissenschaften relativ früh an den Potentialen einer zentralen Verwaltung orientiert. Bereits in den 1960er und -70er Jahren haben sich einige sozialwissenschaftliche Datenzentren zu Verbänden zusammengeschlossen. So stellen das noch heute aktive Inter-University Consortium for Social Political and Social Research (ICPSR)³²¹, das Council of European Social Science Data Archives (CESSDA)³²² und die International Federation of Data Organizations for the Social Sciences

312 <http://www.forschungsdaten.org/informationen-fur-ihre-forschungsvorhaben/disziplin-spezifische-informationen/sozial-und-geisteswissenschaften/>

313 <http://ratswd.de/dat/fdz.php>

314 <http://www.da-ra.de/de/ueber-uns/>

315 Hausstein (2012), S. 6 ff.

316 <http://www.esds.ac.uk/news/newsdetail.asp?id=3268>

317 <http://www.dans.knaw.nl/en/content/categorieen/diensten/persistent-identifiers>

318 <http://www.data-archive.ac.uk/about/projects/store>

319 <http://www.dans.knaw.nl/en/content/categorieen/projecten/driver-ii>

320 Summary of the DANS Strategy Policy 2011-2015, S. 4 (Target 7)

[http://www.dans.knaw.nl/sites/default/files/file/Uitgaven/Strategie/Samenvatting%20strategienota_UK_DEF\(1\).pdf](http://www.dans.knaw.nl/sites/default/files/file/Uitgaven/Strategie/Samenvatting%20strategienota_UK_DEF(1).pdf)

321 <http://www.icpsr.umich.edu/>

322 <http://www.cessda.org/>

(IFDO)³²³ zentrale Anlaufstellen zur Entwicklung und Förderung guter wissenschaftlicher Praxis im Forschungsdatenmanagement dar.

3.5.2. Technische Infrastruktur

Untersuchung der Forschungsdaten-Repositoryen

Die Untersuchung der technischen Infrastruktur in den Sozialwissenschaften erfolgte durch die Auswertung der Funktionalitäten ausgewählter Repositoryen. Als repräsentativ für den Fachbereich wurden fünf Forschungsdaten-Repositoryen ausgewählt, die in mindestens zwei Referenzlisten aufgelistet worden waren: Australian Social Science Data Archive (ASSDA)³²⁴, IQSS Dataverse Network (Dataverse)³²⁵, DANS Electronic Archiving System (EASY)³²⁶, Economic and Social Science Data Service (ESDS)³²⁷ und National Archive of Criminal Justice Data (NACJD)³²⁸. Die Ergebnisse der Auswertung sind in Tabelle 9 zusammengefasst.

Funktionalitäten der Repositoryen	ASSDA	Dataverse	EASY	ESDS	NACJD
Datenformate	PDF, POR, DTA, TAB, XLS, CSV, DIV, DBASE	PDF, POR, DTA, TAB, XLS, DAT, SAS, R, Java	PDF, POR, DTA, Java, XML, TXT, CSV, DIC, FRQ, LAB, RNC	PDF, POR, DTA, TAB, XLS, DAT, SAS, MDB	PDF, POR, DTA, DAT, TXT
Herunterladen möglich	Anmeldung erforderlich; für nicht-Mitglieder kostenpflichtig (1000 AUD / Datensatz)	ohne Anmeldung	Anmeldung erforderlich	Anmeldung erforderlich	Anmeldung erforderlich; alternativ über Google oder Facebook-Account
Anmeldung erfolgreich	nein	-	ja	nein	ja
PID	internes ID	Handle	URN	DOI	DOI
Zitationsvorschlag direkt einbezogen	nein	ja	ja	nein	ja
Version	ja	ja	nein	nein	ja
Methodenbeschreibung	ja	ja	ja	ja	ja

323 <http://ec2-50-17-181-92.compute-1.amazonaws.com/>

324 <http://nesstar.assda.edu.au/webview/>

325 <http://dvn.iq.harvard.edu/dvn/>

326 <https://easy.dans.knaw.nl/ui/browse>

327 <http://www.esds.ac.uk/Lucene/Search.aspx>

328 <http://www.icpsr.umich.edu/icpsrweb/NACJD/das.jsp>

Nutzungsrechte	Antrag muss genehmigt werden; nur akademische Nutzung	Nutzungs- vertrag; nur akademische Nutzung	Nutzungs- vertrag; nur akademische Nutzung	Nutzungs- vertrag; nur akademische Nutzung	Nutzungs- vertrag; nur akademische Nutzung
Verwandte Publikationen	ja	ja	ja	ja	ja
Nutzungsstatistik	nein	ja	ja	nein	nein
Deposit-Prozedur	Formular und Daten werden per Post / Email zugeschickt	Online- Formular	Online- Formular	Formular und Daten werden per Post / Email zugeschickt	Online- Formular

Tabelle 9: Auswertung der Forschungsdaten-Repositoryn in den Sozialwissenschaften

Die untersuchten Repositoryn zeichnen sich vor allem durch eingeschränkte Nutzungsbedingungen aus. So stellen vier von fünf Repositoryn die Forschungsdaten nur nach der Anmeldung zur Verfügung. Dabei konnte die Anmeldung in ASSDA und ESDS nicht erfolgreich abgeschlossen werden, da keine Zugangsdaten von dem Repository mitgeteilt wurden. Daneben sind in allen Repositoryn Forschungsdaten zu finden, die ohne zusätzlicher Genehmigung durch den Daten-Urheber nicht zugänglich sind. Für die Festlegung der Nutzungsbedingungen hat jedes Datenarchiv einen eigenen Nutzungsvertrag. Ausdrückliche Erklärung zu den Prinzipien von Open Access oder Einsatz von CC-Lizenzen sind dabei in keinem von diesen zu treffen.

Im Gegensatz dazu hat die überwiegende Mehrheit der Archive den Forschungsdaten persistente Identifikatoren zugewiesen und den Zitationsvorschlag direkt einbezogen. Auf der Basis der Daten entstandene Publikationen werden in allen Repositoryn aufgelistet und oft direkt verlinkt. Außerdem stellen die Sozialwissenschaften mit drei von fünf Repositoryn die einzige Disziplin in der vorliegenden Arbeit dar, in der die Forschungsdaten mit deren Version versehen worden waren.

Allerdings stellt das von der Harvard University betriebene Dataverse das einzige Repository dar, das über alle für *enhanced publications* notwendige Merkmale bezüglich Referenzierbarkeit und Zugänglichkeit der Forschungsdaten verfügt. Bemerkenswert ist dabei auch, dass solche Schlussfolgerung nicht durch fehlende technische Funktionalitäten bedingt ist, sondern einer organisatorischen Entscheidung bzw. eines Kulturwandels hin zum *data sharing* innerhalb der Disziplin bedarf. Somit könnte z.B. das Aufheben der Zugangseinschränkungen zur Integration von Forschungsdaten erheblich beitragen.

Schließlich ist die Landschaft der disziplinspezifischen Datenzentren in den Sozialwissenschaften oft durch eine nationale Ausrichtung geprägt. Dies kann möglicherweise durch die konsultative Rolle der

Disziplin für Entscheidungen in der Sozialpolitik erklärt werden. Außerdem werden gesellschaftliche Wandlungsprozesse auf internationaler Ebene beobachtet und bedingen somit repräsentative Forschungsstellen in zahlreichen Staaten. Im Zusammenhang mit regierungsweiten Initiativen zu Open Data sollte dadurch der Zugang zu den sozialwissenschaftlichen Daten zukünftig noch mehr an Bedeutung gewinnen.

Beispiele für Forschungsdatenintegration

Im Bereich der Sozialwissenschaften lassen sich bereits einige Versuche zur Integration von Forschungsdaten beobachten. So werden beispielsweise in den Repositorien abgelegte Forschungsdaten oft mit den daraus entstandenen wissenschaftlichen Publikationen verlinkt.³²⁹ Andererseits findet die Verlinkung in der Gegenrichtung - also von der Zeitschrift zum Repository - nur selten statt. Vielmehr sind zahlreiche Beispiele zu finden, bei denen auf die zugrundeliegenden Daten trotz deren öffentlicher Zugänglichkeit nicht hingewiesen wird.³³⁰ Von manchen anderen Zeitschriften werden die Daten zusammen mit dem Artikel zum Download bereitgestellt, auch wenn diese in einem Repository bereits archiviert werden.³³¹ Somit werden die Potentiale aus der Kooperation mit Forschungsdaten-Repositorien nicht ausgenutzt, indem die Daten seitens der Zeitschrift nicht eigenständig referenzierbar sind und dazu doppelt archiviert werden müssen. Außerdem werden solche Datensätze oft nur in Verbindung mit dem Zeitschrift-Abonnement zugänglich gemacht.

Als weiteres Beispiel zur Integration von Forschungsdaten dient der StatLink-Service der OECD. Die Publikationen der OECD werden dabei mittels eines DOI mit den zugrundeliegenden Dokumenten, Daten oder Grafiken verlinkt.³³² Besonders oft findet solche Verlinkung von Abbildungen in den publizierten Berichten mit den zugrundeliegenden aktualisierten Excel-Tabellen statt.³³³

Auf nationaler Ebene ist das von 2011 bis 2013 laufende durch die DFG geförderte Projekt „InFoLiS“ – Integration von Forschungsdaten und Literatur in den Sozialwissenschaften – von besonderem Interesse. Im Rahmen des Projektes sollen die Forschungsdaten im Datenarchiv von GESIS und die Literatur im Katalog der Universitätsbibliothek Mannheim gegenseitig vernetzt werden. Das bestmögliche Verfahren soll dabei durch die automatisierte Extraktion von Referenzen gefunden

329 Siehe z.B. <http://hdl.handle.net/1902.1/10098>

330 Siehe z.B. <http://dx.doi.org/10.1111/j.1468-2478.2005.00370.x>

331 Siehe z.B. <http://dx.doi.org/10.1177/0022343304047432> (Journal of Peace Research), <http://hdl.handle.net/1902.1/10219> (Dataverse)

332 <http://www.oecd.org/about/publishing/aboutoecdstatlink.htm>

333 Vgl. Hinweise zu StatLink bei den Tabellen im PDF-Dokument <http://dx.doi.org/10.1787/oif-2009-1-en>

werden.³³⁴ Die Ergebnisse des Projekts werden dabei zur zukünftigen die Umsetzung von *enhanced publications* in den Sozialwissenschaften beitragen.

3.5.3. Politische Rahmenbedingungen

Die Förderung in den Sozialwissenschaften erfolgt durch ihre Zuordnung zu den angrenzenden Fachbereichen. Die politischen Rahmenbedingungen der Disziplin wurden nach den Vorgaben von drei Förderorganisationen untersucht: die „Research Data Archiving Policy“ des Social Sciences and Humanities Research Council of Canada (SSHRC)³³⁵ in Kanada, die „Research Data Policy“ des Economic and Social Research Council (ESRC)³³⁶ in Großbritannien und den „Data Management for NSF SBE Directorate Proposals and Awards“ des Social, Behavioral and Economic Sciences Directorate der NSF (NSF-SBE)³³⁷ in den USA. Die Ergebnisse der Auswertung sind in Tabelle 10 zusammengefasst.

Anforderungen der Förderorganisation	SSHRC	ESRC	NSF-SBE
	Canada	UK	USA
Forschungsdaten müssen öffentlich zugänglich gemacht werden ab	k. A.	Frühling 2011	Januar 2011
Zeitraumen	innerhalb von 2 Jahren nach Projektende	innerhalb von 3 Monaten nach Förderungsende	zum frühestmöglichen Zeitpunkt
Datenmanagementplan erforderlich	nein	ja	ja
Ausnahmen erlaubt	vertrauliche Daten	sensible Daten	vertrauliche oder geschützte Daten
Zusätzliche Anforderungen	k. A.	für die Nachnutzung notwendige Dokumentation beifügen	festgelegt in der jeweiligen Ausschreibung
Vorgegebenes Repository	Bibliothek oder Datenservice der Forschungsorganisation	ESDS (UKDA)	keine Bevorzugung
Unterstützung	k. A.	eigenes Datenservice, Personal und Anleitungen	Universitätseinrichtungen oder Fachorganisationen
Sanktionen / Kontrolle	k. A.	Abschlusszahlung kann vorenthalten werden	Kurzbericht in den nachstehenden Antragstellungen

Tabelle 10: Anforderungen der Förderorganisationen in den Sozialwissenschaften

334 <http://www.bib.uni-mannheim.de/1158.html>

335 http://www.sshrc-crsh.gc.ca/about-au_sujet/policies-politiques/statements-enonces/edata-donnees_electroniques-eng.aspx

336 http://www.esrc.ac.uk/_images/Research_Data_Policy_2010_tcm8-4595.pdf

337 http://www.nsf.gov/sbe/SBE_DataMgmtPlanPolicy.pdf

Alle untersuchten Förderorganisationen fordern die Forscher grundsätzlich dazu auf, ihre Forschungsergebnisse so früh wie möglich zugänglich zu machen. Dabei gestattet jede einzelne Organisation ausdrücklich solche Ausnahmen, in denen aufgrund des Datenschutzes oder der Forschungsethik die produzierten Datensätze nicht öffentlich zugänglich gemacht werden dürfen. In solchen Fällen werden die Forscher gebeten, diese Umstände rechtzeitig der Organisation mitzuteilen. Auch bei keiner Veröffentlichung der Daten sind die Forscher jedoch verpflichtet diese zuverlässig zu archivieren. Insgesamt weisen die Richtlinien von ESRC die strengsten Vorgaben auf. So sind die Forscher verpflichtet innerhalb eines vergleichsweise kurzen Zeitrahmens ihre Forschungsdaten in einem vom ESRC vorgeschriebenen Repository abzulegen. Außerdem kann die Abschlusszahlung bei Nicht-Einhaltung mit den Vorgaben einbehalten werden. Andere Förderorganisationen bevorzugen kein konkretes Repository und verweisen stattdessen auf die Unterstützung durch die Forschungseinrichtungen, insbesondere durch die Hochschulen. Eine Verpflichtung zur kurzen Darstellung des Datenmanagements bei späteren Antragstellungen hat zwar auch das NSF-SBE, die Einstellung der Finanzierung wird aber nicht ausdrücklich erwähnt. Die strikte Position des ESRC könnte dadurch erklärt werden, dass alle für die Erfüllung der Voraussetzungen notwendige Einrichtungen vom Council selbst bereitgestellt werden.

4. Fazit

Die vorliegende Masterarbeit hatte zum Ziel, die Bedingungen für die Umsetzung von *enhanced publications* in ausgewählten Wissenschaftsdisziplinen zu untersuchen. Wie bereits von einigen Quellen identifiziert, sind die Unterschiede im Umgang mit Forschungsdaten innerhalb der Disziplinen teilweise so groß, dass die *domains* oder sogar spezialisierte Forschergruppen analysiert werden sollten. Diese Erkenntnis wurde bei der Auswahl der zu untersuchenden Fachbereiche berücksichtigt. So stellt die Klimaforschung eine Unterdisziplin der Geowissenschaften, die Kristallographie eine solche der Chemie und die Ökologie eine Unterdisziplin der Biologie dar. Die Wissenschaftsdisziplin der Archäologie wird dabei oft dem Großbereich der Geisteswissenschaften untergeordnet, konnte aber als ein äußerst interdisziplinärer Fachbereich im Kontext der Untersuchung betrachtet werden. Im Vergleich dazu, stellen die Sozialwissenschaften eher einen Großbereich selbst dar und wurden mehr spezifisch als empirisch arbeitende Sozialwissenschaften analysiert.

Der Umgang mit Forschungsdaten in ausgewählten Disziplinen wurde hinsichtlich dreier Teilaspekte untersucht: Organisation und Arbeitsweise der Disziplin, technische Infrastruktur und politische Rahmenbedingungen. Es wurde davon ausgegangen, dass bei positiven Ergebnissen in allen drei Teilen sich *enhanced publications* besser umsetzen lassen. Während aus organisatorischer Sicht alle untersuchten Disziplinen über spezielle Metadatenschemata für die Erschließung von Forschungsdaten und koordinierende Organisationen verfügten, ließen sich große Unterschiede durch die Untersuchung der Funktionalitäten in Forschungsdaten-Repositoryn und der Vorgaben der Förderorganisationen in der jeweiligen Disziplin aufzeigen.

Bei der Untersuchung der Forschungsdaten-Repositoryn wurden kritische und motivierende Aspekte für *enhanced publications* definiert. So wurde die Zugänglichkeit von Forschungsdaten, deren persistente Identifizierung und die Beschreibung der Datenerhebungsmethoden als notwendig angesehen. Gleichzeitig dienten eine direkte Einbindung des Zitationsvorschlages, Hinweise auf aus den Daten entstandene Publikationen und Nutzungsstatistiken als zusätzlichen Funktionalitäten, die die Autoren zur Bereitstellung der Daten motivieren könnten. Außerdem wurden auch Deposit-Prozeduren der Repositoryn betrachtet, um die tatsächlichen *data sharing*-Möglichkeiten zu veranschaulichen.

Eine offene *data sharing*-Kultur war insbesondere in der Archäologie, Ökologie und Kristallographie deutlich erkennbar. So konnten die Forschungsdaten aus vier von fünf untersuchten Repositoryn jeweils ohne Anmeldung heruntergeladen werden. Außerdem waren oft explizite Aussagen zu Open Access, also freier Nutzung von Daten, und der Einsatz von CC-Lizenzen zu finden. Ähnliche Nutzungsbestimmungen wiesen die Forschungsdaten-Repositoryn in der Klimaforschung auf, indem

sogar zwei davon ausdrücklich Unterstützung von Open Data-Prinzipien zeigten. Allerdings ließen die klimawissenschaftlichen Repositorien differenziertere Zugangsbestimmungen feststellen: zum Herunterladen von Daten war in manchen davon die Anmeldung erwünscht oder erforderlich. Ein Repository war dabei nur für Partnerorganisationen zugänglich.

Am restriktivsten sind die Nutzungsbedingungen in den sozialwissenschaftlichen Forschungsdaten-Repositorien aufgefunden worden. In nur einem von fünf untersuchten Repositorien konnten die Daten ohne vorherige Anmeldung heruntergeladen werden. In zwei anderen Repositorien konnte die Anmeldung durch Ausfüllen eines Online-Formulars erfolgreich abgeschlossen werden. Durch die restlichen zwei Repositorien wurde keine Berechtigung erteilt und die Forschungsdaten blieben somit unzugänglich. Zusätzlich werden in einem davon auch Gebühren für jeden Datensatz erhoben. Auch bezüglich der Nutzungsrechte erwiesen sich die untersuchten Repositorien vergleichsweise einschränkend: jedes Repository hatte einen eigenen Nutzungsvertrag, bei einem davon musste das Antragsformular zusätzlich genehmigt werden.

Des Weiteren wurde als notwendig für die Integration von Forschungsdaten in *enhanced publications* eindeutige persistente Identifizierung vorausgesetzt. In diesem Bereich wiesen die sozialwissenschaftlichen Repositorien einen fortgeschrittenen Zustand auf. So wurden in vier von fünf Repositorien die Daten mit persistenten Identifikatoren versehen. In der Archäologie, Klimaforschung und Ökologie waren es jeweils zwei, gefolgt von der Kristallographie mit nur einem Repository. Ebenso für die Nachnutzung und Nachprüfbarkeit der integrierten Daten wurde die Methodenbeschreibung zu deren Gewinnung vorausgesetzt. Dies stellte eher eine Regel in allen untersuchten Disziplinen dar, indem nur ein Repository in der Archäologie und eines in der Klimaforschung keine nähere Informationen zu der Datenerhebung zur Verfügung gestellt haben.

Hinsichtlich der kritischen Aspekte für die Integration von Forschungsdaten beim wissenschaftlichen Publizieren konnte keine der untersuchten Disziplinen positive Ergebnisse in allen drei Teilbereichen aufweisen. Allerdings werden z.B. statt persistenten Identifikatoren meist eigene interne in jeder Datenbank eingesetzt und erlauben somit eine direkte Verlinkung. Aus dieser Sicht kann die freie Zugänglichkeit von Forschungsdaten als das allerwichtigste Kriterium angesehen und die Bedingungen in der Archäologie, Ökologie und Kristallographie damit am positivsten bewertet werden.

Des Weiteren wurden motivierende Aspekte für das *data sharing* und die Einbindung von Forschungsdaten in die Publikationen untersucht. So zeigte die Auswertung nach der Einbindung eines direkten Zitationsvorschlages und dem Hinweis auf verwandte Publikationen eine gleichmäßige Verteilung. Die meisten Repositorien mit diesen Funktionalitäten waren in der Archäologie und

Klimaforschung zu finden, gefolgt von Sozialwissenschaften und Kristallographie, am wenigsten in der Ökologie. Die Bereitstellung der Nutzungsstatistiken scheint dabei noch nicht so üblich zu sein. So waren die Download-Zahlen in der Archäologie und in den Sozialwissenschaften jeweils in zwei Repositorien zu finden, gefolgt von der Ökologie mit einem Repository. Die untersuchten Repositorien in der Kristallographie und Klimaforschung stellten keine Nutzungsstatistiken zur Verfügung. Nach der Auswertung der motivationssteigernden Funktionalitäten zeichnen sich somit die Repositorien in der Archäologie und in den Sozialwissenschaften aus.

Schließlich wurden die Vorgaben der Förderorganisationen der jeweiligen Wissenschaftsdisziplin ausgewertet. Insbesondere wurden dabei die Forderung zur Einreichung eines Datenmanagementplans sowie zusätzliche durch jeweilige Organisation definierte Anforderungen und die Kontrolle bei der Erfüllung der Vorgaben betrachtet. Die Forderung nach der Einreichung des Datenmanagementplans unterscheidet sich unwesentlich über die Disziplinen. Bis auf eine Organisation in der Kristallographie war eine verbindliche Forderung in zwei von drei Richtlinien der jeweiligen Disziplin zu finden. Hinsichtlich zusätzlicher Anforderungen und Kontrollen erwiesen sich die Bestimmungen in der Klimaforschung und der Ökologie am strengsten. Das Einstellen der Zahlungen bis zur Erfüllung der Vorgaben waren jeweils einmal in den Vorgaben für Projekte in der Klimaforschung und den Sozialwissenschaften zu finden. Allerdings waren solche Sanktionen in beiden Fällen in Großbritannien zu finden und würden keine direkte Auswirkung auf Forschungsprojekte in anderen Ländern haben.

Außerdem lassen sich die Ausnahmen bei den Vorgaben der Förderorganisationen zum großen Teil als disziplinübergreifend feststellen. So werden insbesondere personenbezogene, vertrauliche, kommerzielle und für die Nachnutzung nicht geeignete Daten von den Anforderungen ausgeschlossen. Zusätzlich werden manche disziplinspezifische Daten sehr sensibel behandelt, z.B. das größere Ausmaß der personenbezogenen Daten in den Sozialwissenschaften oder präzise Koordinaten der Ausgrabungsstätten in der Archäologie. Solche und ähnliche Einzelheiten müssen von den Repositorien und demzufolge auch bei der Integration von Forschungsdaten mitberücksichtigt werden.

Abschließend lassen sich nach der Gesamteinschätzung die besten Bedingungen für die Umsetzung von *enhanced publications* in der Archäologie feststellen. In den anderen untersuchten Disziplinen ergibt sich kein einheitliches Bild, da die Auswertung der Teilaspekte sehr unterschiedlich ausfällt. Allerdings sollen diese Ergebnisse als eine Bestandsaufnahme nach den bestimmten Kriterien betrachtet werden. Manche Repositorien befanden sich zum Untersuchungszeitpunkt noch im Ausbau, ebenso wie die Richtlinien mancher Förderorganisationen. Im Allgemeinen lässt sich somit ein positives Klima für die Einbindung von Forschungsdaten beim elektronischen Publizieren feststellen. Im Kontext der weltweiten Initiativen ist eine weitere positive Entwicklung zu erwarten.

5. Literaturverzeichnis

- Allianz der deutschen Wissenschaftsorganisationen (2010): Grundsätze zum Umgang mit Forschungsdaten. Online verfügbar unter <http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaeetze/>, zuletzt geprüft am 29.01.2012.
- Bailey, Charles W. (1994): Scholarly electronic publishing on the internet, the NREN, and the NII: Charting possible futures. In: *Serials Review* 20 (3), S. 7–16. DOI: 10.1016/0098-7913(94)90034-5.
- Bechhofer, Sean; Buchan, Iain; Roure, David de; Missier, Paolo; Ainsworth, John; Bhagat, Jiten et al. (2011): Why linked data is not enough for scientists. In: *Future Generation Computer Systems*. DOI: 10.1016/j.future.2011.08.004.
- Bellini, Emanuele (2012): APARSEN Deliverable D22.1 Persistent Identifiers Interoperability Framework. Online verfügbar unter http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/04/APARSEN-REP-D22_1-01-1_9.pdf.
- Biegon, Dagmar (2006): Project StORe: Biosciences interviews. Online verfügbar unter <http://hdl.handle.net/1842/1460>.
- Bijsterbosch, Magchiel; Elbaek, Mikael Karstens; Hochstenbach, Patrick; Ludwig, Jens; Schmeltz Pedersen, Gert; Russell, Rosemary et al. (2008): Technology Watch report. DRIVER II Deliverable D4.3. Hg. v. Karen van Godtsenhoven. Online verfügbar unter http://www.driver-repository.eu/component/option,com_jdownloads/Itemid,58/task,view.download/cid,50/, zuletzt geprüft am 03.02.2012.
- Blake, Judith A.; Bult, Carol J. (2006): Beyond the data deluge: Data integration and bio-ontologies. In: *Journal of Biomedical Informatics* 39 (3), S. 314–320. DOI: 10.1016/j.jbi.2006.01.003.
- Borgman, Christine L. (2008): Data, disciplines, and scholarly publishing. In: *Learn. Pub* 21 (1), S. 29–38. DOI: 10.1087/095315108X254476.
- Burton, Guy (2006): Project StORe: Social Science report. Online verfügbar unter <http://hdl.handle.net/1842/1418>.
- Büttner, Stephan; Hobohm, Hans-Christoph; Müller, Lars (Hg.) (2011): Handbuch Forschungsdatenmanagement. Bad Honnef: Bock + Herchen. Online verfügbar unter <http://www.forschungsdatenmanagement.de/>.
- Callaghan, Sarah (2009): Overlay Journal Infrastructure for Meteorological Sciences (OJIMS). Final report. Online verfügbar unter <http://repository.jisc.ac.uk/id/eprint/349>.
- Chavan, Vishwas; Penev, Lyubomir (2011): The data paper: a mechanism to incentivize data publishing in biodiversity science. In: *BMC Bioinformatics* 12 (Suppl 15), S. S2. DOI: 10.1186/1471-2105-12-S15-S2.
- Cheung, Kwok; Lashtabeg, Anna; Drennan, John; Hunter, Jane (2008): SCOPE: A Scientific Compound Object Publishing and Editing System. In: *International Journal of Digital Curation* 3 (2), S. 4–18. Online verfügbar unter <http://www.ijdc.net/index.php/ijdc/article/view/84/55>.
- Dally, Ortwin; Fless, Friederike; Förtsch, Reinhard (2012): Altertumswissenschaften. In: Heike Neuroth, Stefan Strathmann, Achim Oßwald, Regine Scheffel, Jens Klump und Jens Ludwig (Hg.): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg, Göttingen: Hülsbusch; Univ.-Verl. Göttingen, S. 161–178.
- Dallmeier-Tiessen, Sünje (2011): Strategien bei der Veröffentlichung von Forschungsdaten. In: Stephan Büttner, Hans-Christoph Hobohm und Lars Müller (Hg.): Handbuch Forschungsdatenmanagement. Bad Honnef: Bock + Herchen, S. 157–168. Online verfügbar unter http://opus.kobv.de/fhpotdamsam/volltexte/2011/237/pdf/3.2_Strategien_bei_der_VerAeffentlichung_von_Forschungsdaten_.pdf.
- Deutsche Forschungsgemeinschaft (1998): Vorschläge zur Sicherung guter wissenschaftlicher Praxis. Empfehlungen der Kommission "Selbstkontrolle in der Wissenschaft". Weinheim. Online verfügbar unter http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf.

- Deutsche Initiative für Netzwerkinformation e.V. (2009): Positionspapier Forschungsdaten. Arbeitsgruppe "Elektronisches Publizieren". Online verfügbar unter <http://edoc.hu-berlin.de/series/dini-schriften/2009-10/PDF/10.pdf>, zuletzt geprüft am 25.02.2012.
- Fielding, Nigel; Lee, Raymond M.; Blank, Grant (Hg.) (2008): The SAGE handbook of Internet of online research methods. Los Angeles ; London: SAGE.
- Friebertshäuser, Barbara (Hg.) (2010): Handbuch qualitative Forschungsmethoden in der Erziehungswissenschaft. 3. Aufl. Weinheim, München: Juventa-Verl.
- Fritzsche, Bernadette (2012): Klimaforschung. In: Heike Neuroth, Stefan Strathmann, Achim Oßwald, Regine Scheffel, Jens Klump und Jens Ludwig (Hg.): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg, Göttingen: Hülbusch; Univ.-Verl. Göttingen, S. 195–212.
- Garfield, Eugene (2005): The Agony and The Ecstasy – The History and Meaning of the Journal Impact Factor. International Congress on Peer Review And Biomedical Publication. Online verfügbar unter <http://garfield.library.upenn.edu/papers/jifchicago2005.pdf>.
- Giaretta, David; Hudson, Richard L. (Hg.) (2010): Riding the wave: How Europe can gain from the rising tide of scientific data. Final report of the High Level Expert Group on Scientific Data. A submission to the European Commission. Unter Mitarbeit von John Wood, Thomas Andersson, Achim Bachem, Christoph Best, Françoise Genova, Diego R. Lopez et al. Online verfügbar unter <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>
- Grazulis, S.; Daskevicius, A.; Merkys, A.; Chateigner, D.; Lutterotti, L.; Quiros, M. et al. (2011): Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. In: *Nucleic Acids Research* 40 (D1), S. D420. DOI: 10.1093/nar/gkr900.
- Groth, Paul; Gibson, Andrew; Velterop, Jan (2010): The anatomy of a nanopublication. In: *Information Services and Use* 30 (1), S. 51–56. DOI: 10.3233/ISU-2010-0613. Online verfügbar unter <http://iospress.metapress.com/content/ftkh21q50t521wm2/>.
- Hamann, Ilse (2011): Climate Research. Chapter F. In: Christian Meier zu Verl und Wolfram Horstmann (Hg.): Studies on subject-specific requirements for open access infrastructure. 1. Aufl. Bielefeld: Univ.-Bibliothek, S. 215–310. DOI: 10.2390/PUB-2011-8.
- Hansen, Sue Ellen; Iverson, Jeremy; Jensen, Uwe; Orten, Hilde; Vompras, Johanna (2010): Enabling Longitudinal Data Comparison Using DDI. In: *DDI Alliance Working Papers Series* (2). DOI: 10.3886/DDIlongitudinal02.
- Harnad, Stevan (1992): Interactive publication: Extending the American Physical Society's discipline-specific model for electronic publishing. In: *Serials Review* 18 (1-2), S. 58–61. DOI: 10.1016/0098-7913(92)90044-W.
- Hausstein, Brigitte (2012): Die Vergabe von DOI-Namen für Sozial- und Wirtschaftsdaten. Serviceleistungen der Registrierungsagentur da|ra (RatSWD Working Paper Series, Nr. 193). Online verfügbar unter http://www.ratswd.de/download/RatSWD_WP_2012/RatSWD_WP_193.pdf.
- Hausstein, Brigitte; Zenk-Möltgen, Wolfgang; Wilde, Anja; Schleinstein, Natalija (2011): da|ra Metadatenschema, Version 1.0. GESIS Working Papers 2011/14.
- Hense, Andreas; Quadt, Florian (2011): Acquiring High Quality Research Data. In: *D-Lib Magazine* 17 (1/2). DOI: 10.1045/january2011-hense.
- Hogenaar, Arjan; Tjalsma, Heiko; Priddy, Mike (2011): Research in the Humanities and Social Sciences. Chapter E. In: Christian Meier zu Verl und Wolfram Horstmann (Hg.): Studies on subject-specific requirements for open access infrastructure. 1. Aufl. Bielefeld: Univ.-Bibliothek, S. 165–214. DOI: 10.2390/PUB-2011-7.
- Hoogerwerf, Maarten (2009): Durable Enhanced Publications (Proceedings of African Digital Scholarship & Curation). Online verfügbar unter http://www.ais.up.ac.za/digi/docs/hoogerwerf_paper.pdf, zuletzt geprüft am 22.07.2012.
- Hull, Daniel (2006): Project StORe: Archaeology report. Online verfügbar unter <http://hdl.handle.net/1842/1413>.
- Hunter, Jane (2006): Scientific Publication Packages – A Selective Approach to the Communication and Archival of Scientific Output. In: *International Journal of Digital Curation* 1 (1), S. 33–52. Online verfügbar unter <http://www.ijdc.net/index.php/ijdc/article/view/8>.
- Huschka, Denis; Oellers, Claudia; Ott, Notburga; Wagner, Gert G. (2011): Datenmanagement und Data Sharing: Erfahrungen in den Sozial- und Wirtschaftswissenschaften. In: Stephan Büttner, Hans-Christoph Hobohm und Lars Müller (Hg.): Handbuch Forschungsdatenmanagement. Bad Honnef: Bock + Herchen, S. 35–48. Online

verfügbar unter

http://opus.kobv.de/fhpotsdam/volltexte/2011/227/pdf/1.3_Datenmanagement_und_Data_Sharing.pdf.

ICPSR (2012): Guide to social science data preparation and archiving. Best practice throughout the data life cycle. [Ann Arbor, Michigan]. Online verfügbar unter <http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf>.

IfA (2008): Standard and guidance for archaeological excavation. Online verfügbar unter http://www.archaeologists.net/sites/default/files/node-files/ifa_standards_excavation.pdf.

Jantz, Ronald (2003): Public Opinion Polls and Digital Preservation. In: *D-Lib Magazine* 9 (11). DOI: 10.1045/november2003-jantz.

Kauppinen, Tomi; Espindola, Giovana Mira de (2011): Linked Open Science-Communicating, Sharing and Evaluating Data, Methods and Results for Executable Papers. In: *Procedia Computer Science* 4, S. 726–731. DOI: 10.1016/j.procs.2011.04.076. Online verfügbar unter <http://www.sciencedirect.com/science/article/pii/S1877050911001347>.

Kauppinen, Tomi; Keßler, Carsten; Pouchard, Line C. (Hg.) (2011): LISC2011. Proceedings of the First International Workshop on Linked Science 2011. In conjunction with the International Semantic Web Conference (ISWC2011). Bonn, Germany, October 24. Online verfügbar unter <http://www.ceur-ws.org/Vol-783/>.

Key Perspectives Ltd. (2010): Data Dimensions: Disciplinary Differences in Research Data Sharing, Reuse and Long term Viability. DCC SCARP Synthesis Report. Digital Curation Centre (DCC). Online verfügbar unter http://www.dcc.ac.uk/sites/default/files/SCARP%20SYNTHESIS_FINAL.pdf.

Kircz, Joost G. (1998): Modularity: the next form of scientific information presentation? In: *Journal of Documentation* 54 (2), S. 210–235. DOI: 10.1108/EUM0000000007185. Online verfügbar unter www.emeraldinsight.com/journals.htm?articleid=864073.

Kircz, Joost G. (2001): New practices for electronic publishing 1: Will the scientific paper keep its form? In: *Learned Publishing* 14 (4), S. 265–272. DOI: 10.1087/095315101753141365.

Kircz, Joost G. (2002): New practices for electronic publishing 2: New forms of the scientific paper. In: *Learned Publishing* 15 (1), S. 27–32. DOI: 10.1087/095315102753303652.

Kommission Zukunft der Informationsinfrastruktur (2011): Gesamtkonzept für die Informationsinfrastruktur in Deutschland. Empfehlungen der Kommission Zukunft der Informationsinfrastruktur im Auftrag der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder. Online verfügbar unter <http://www.leibniz-gemeinschaft.de/?nid=infrastr>.

Kuipers, Tom; van der Hoeven, Jeffrey (2009): PARSE.Insight: INSIGHT into issues of Permanent Access to the Records of Science in Europe. Deliverable D3.4. Online verfügbar unter http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf.

Lautenschlager, Michael (2011): Institutionalisierte „Data Curation Services“. In: Stephan Büttner, Hans-Christoph Hobohm und Lars Müller (Hg.): *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock + Herchen, S. 149–156. Online verfügbar unter http://opus.kobv.de/fhpotsdam/volltexte/2011/236/pdf/3.1_Institutionalisierte_Data_Curation_Services.pdf.

Lynch, Clifford (2007): The Shape of the Scientific Article in The Developing Cyberinfrastructure. In: *CTWatch Quarterly* 3 (3). Online verfügbar unter <http://www.ctwatch.org/quarterly/articles/2007/08/the-shape-of-the-scientific-article-in-the-developing-cyberinfrastructure/>.

Lyon, Liz; Rusbridge, Chris; Neilson, Colin; Whyte, Angus (2010): DCC SCARP: Disciplinary Approaches to Sharing, Curation, Reuse and Preservation. JISC Final Report. JISC; Digital Curation Centre (DCC). Online verfügbar unter <http://www.dcc.ac.uk/sites/default/files/documents/scarp/SCARP-FinalReport-Final-SENT.pdf>.

Martín, Elizabeth; Ballard, Grant (2010): Data Management Best Practices and Standards for Biodiversity Data Applicable to Bird Monitoring Data. North American Bird Conservation Initiative, Monitoring Subcommittee. Online verfügbar unter <http://www.nabci-us.org/aboutnabci/bestdatamanagementpractices.pdf>.

Meier zu Verl, Christian; Horstmann, Wolfram (Hg.) (2011): Studies on subject-specific requirements for open access infrastructure. 1. Aufl. Bielefeld: Univ.-Bibliothek.

Meier zu Verl, Christian; Horstmann, Wolfram (2011): Subject-Specific Requirements for Open Access Infrastructure - Attempt at a Synthesis. Chapter H. In: Christian Meier zu Verl und Wolfram Horstmann (Hg.):

Studies on subject-specific requirements for open access infrastructure. 1. Aufl. Bielefeld: Univ.-Bibliothek, S. 359–381. DOI: 10.2390/PUB-2011-3.

Mons, Barend; Velterop, Jan: Nano-Publication in the e-science era. Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009), Washington, DC. Online verfügbar unter [http://www.w3.org/wiki/images/4/4a/HCLSS\\$ISWC2009\\$Workshop\\$Mons.pdf](http://www.w3.org/wiki/images/4/4a/HCLSS$ISWC2009$Workshop$Mons.pdf).

Mornati, Susanna; Hedlund, Turid (Hg.) (2009): Rethinking Electronic Publishing : ELPUB 2009. ELPUB 2009. Milan, 10-12 June 2009: Nuova Cultura.

Murray-Rust, Peter (2007): Data-driven - science a scientist's view. Position paper at NSF/JISC Repositories Workshop on Data-Driven Science & Scholarship, April 17-19, 2007, Phoenix, Arizona. Online verfügbar unter <http://www.sis.pitt.edu/~repwksch/papers/murray.pdf>, zuletzt geprüft am 03.03.2012.

Neuroth, Heike; Strathmann, Stefan; Oßwald, Achim; Scheffel, Regine; Klump, Jens; Ludwig, Jens (Hg.) (2012): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg, Göttingen: Hülsbusch; Univ.-Verl. Göttingen. Online verfügbar unter http://www.nestor.sub.uni-goettingen.de/bestandsaufnahme/nestor_lza_forschungsdaten_bestandsaufnahme.pdf

NHMRC (2007): Australian Code for the Responsible Conduct of Research. Canberra. Online verfügbar unter http://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/r39.pdf.

Nieschulze, Jens; König-Ries, Birgitta (2012): Funktionelle Biodiversität. In: Heike Neuroth, Stefan Strathmann, Achim Oßwald, Regine Scheffel, Jens Klump und Jens Ludwig (Hg.): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg, Göttingen: Hülsbusch; Univ.-Verl. Göttingen, S. 213–226.

NWO (2010): The open access policy of NWO in brief. Online verfügbar unter [http://www.nwo.nl/files.nsf/pages/NWOP_8BWJML_Eng/\\$file/Flyer%20Open%20Access%20NWO.pdf](http://www.nwo.nl/files.nsf/pages/NWOP_8BWJML_Eng/$file/Flyer%20Open%20Access%20NWO.pdf)

OECD (2007): Principles and Guidelines for Access to Research Data from Public Funding. OECD Publishing. Online verfügbar unter <http://www.oecd.org/science/scienceandtechnologypolicy/38500813.pdf>

Penev, Lyubomir; Mietchen, Daniel; Chavan, Vishwas; Hagedorn, Gregor; Remsen, David; Smith, Vincent; Shotton, David (2011): Pensoft Data Publishing Policies and Guidelines for Biodiversity Data. Online verfügbar unter http://www.pensoft.net/J_FILES/Pensoft_Data_Publishing_Policies_and_Guidelines.pdf.

Piwowar, Heather A.; Day, Roger S.; Fridsma, Douglas B.; Ioannidis, John (2007): Sharing Detailed Research Data Is Associated with Increased Citation Rate. In: *PLoS ONE* 2 (3), S. e308. DOI: 10.1371/journal.pone.0000308.

Polydoratou, Panayiota (2006): Project StORe: Chemistry Report. Online verfügbar unter <http://hdl.handle.net/1842/1415>.

PRC (2010): Access vs. Importance. A global study assessing the importance of and ease of access to professional and academic information. Phase I Results. Online verfügbar unter <http://www.publishingresearch.net/documents/PRCAccessvsImportanceGlobalstudyPhaseIOct2010.pdf>.

Pryor, Graham (2007): Project StORe: Work Package 5 evaluation. Online verfügbar unter <http://hdl.handle.net/1842/2012>.

Rasmussen, Karsten Boye (2008): General Approaches to Data Quality and Internet-generated Data. In: Nigel Fielding, Raymond M. Lee und Grant Blank (Hg.): The SAGE handbook of Internet of online research methods. Los Angeles ; London: SAGE, S. 79–96.

Reilly, Susan; Schallier, Wouter; Schrimpf, Sabine; Smit, Eefke; Wilkinson, Max (2011): ODE Report on Integration of Data and Publications. Opportunities for Data Exchange (ODE) project.

Rzepa, Henry S.; Murray-Rust, Peter (2001): A new publishing paradigm: STM articles as part of the semantic web. In: *Learned Publishing* 14 (3), S. 177–182. DOI: 10.1087/095315101750240421.

Schäfer, Felix; Thänert, Sabine: Aktuelle Entwicklungen am Deutschen Archäologischen Institut im Umgang mit altertumswissenschaftlichen Forschungsdaten. RatSWD WP 178. Online verfügbar unter http://www.ratswd.de/download/RatSWD_WP_2011/RatSWD_WP_178.pdf.

Schriger, David L.; Chehrazi, Ariana C.; Merchant, Rashida M.; Altman, Douglas G. (2011): Use of the Internet by Print Medical Journals in 2003 to 2009: A Longitudinal Observational Study. In: *Annals of Emergency Medicine* 57 (2), S. 153–160.e3. DOI: 10.1016/j.annemergmed.2010.10.008.

Shotton, David (2009): Semantic publishing: the coming revolution in scientific journal publishing. In: *Learn. Pub* 22 (2), S. 85–94. DOI: 10.1087/2009202.

Tenopir, Carol; Allard, Suzie; Douglass, Kimberly; Aydinoglu, Arsev Umur; Wu, Lei; Read, Eleanor et al. (2011): Data Sharing by Scientists: Practices and Perceptions. In: *PLoS ONE* 6 (6), S. e21101. DOI: 10.1371/journal.pone.0021101.

TIB; FIZ Chemie; Universität Paderborn (2010): Konzeptstudie Vernetzte Primärdaten-Infrastruktur für den Wissenschaftler-Arbeitsplatz in der Chemie. Online verfügbar unter http://www.tib-hannover.de/fileadmin/projekte/primaer-chemie/Konzeptstudie_Forschungsdaten_Chemie.pdf.

van Bentum, Maarten; Hogenaar, Arjan; Vierkant, Dennis: Enhanced Scientific Communication by Aggregated Publications Environments (ESCAPE). In: Susanna Mornati und Turid Hedlund (Hg.): *Rethinking Electronic Publishing* : ELPUB 2009. ELPUB 2009. Milan, 10-12 June 2009: Nuova Cultura, S. 601–606. Online verfügbar unter <http://conferences.aepic.it/index.php/elpub/elpub2009/paper/view/108/48>.

Verhaar, Peter (2008): Report on Object Models and Functionalities. DRIVER II Deliverable D4.2. Online verfügbar unter http://www.driver-repository.eu/component/option,com_jdownloads/Itemid,58/task,view.download/cid,54/, zuletzt geprüft am 03.02.2012.

Wissenschaftsrat (2011): Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften. Berlin (Drs. 10465-11). Online verfügbar unter <http://www.wissenschaftsrat.de/download/archiv/10465-11.pdf>.

WMO (1988): General Meteorological Standards and Recommended Practices. Technical Regulations. No. 49, Volume I. Online verfügbar unter http://www.wmo.int/pages/prog/www/WIS/Publications/49%20%28Technical%20Regulations%29/49%20-%20Volume%20I/49_I_E.pdf

Woutersen-Windhower, Saskia; Vernooy-Gerritsen, Marjan (2009): Enhanced publications. Linking publications and research data in digital repositories. Amsterdam: Amsterdam University Press.

Woutersen-Windhower, Saskia; Brandsma, Renze (2009): Report on Enhanced Publications state-of-the-art. DRIVER II Deliverable D4.1. Online verfügbar unter http://www.driver-repository.eu/component/option,com_jdownloads/Itemid,83/task,view.download/cid,53/, zuletzt geprüft am 03.02.2012.

Zitierte Forschungsdaten

European Commission (2012): Eurobarometer 75.1 (2011). TNS OPINION & SOCIAL, Brussels. GESIS Data Archive, Cologne. Study ZA5479, Data file Version 5.0.0, <http://dx.doi.org/10.4232/1.11413>

OECD (2009), Table A5.1b. Enjoyment of science, in: *Education at a Glance*, OECD Publishing. http://dx.doi.org/10.1787/eag-2009-tablea5_1b-en

Piowar et al. (2007) Daten aus: Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLoS ONE* 2(3). Figure 1: <http://dx.doi.org/10.1371/journal.pone.0000308.g001>, Figure 2 <http://dx.doi.org/10.1371/journal.pone.0000308.g002>

Soil Systems, Inc.; Hill, Rebecca (2011) Pueblo Grande DMB Unknown BG demography & paleopathology dataset. (tDAR ID: 367878) <http://dx.doi.org/10.6067/XCV8N87870>

6. Tabellenverzeichnis

Tabelle 1: Auswertung der Forschungsdaten-Repositories in der Archäologie	22
Tabelle 2: Anforderungen der Förderorganisationen in der Archäologie	25
Tabelle 3: Auswertung der Forschungsdaten-Repositories in der Klimaforschung	30
Tabelle 4: Anforderungen der Förderorganisationen in der Klimaforschung	33
Tabelle 5: Auswertung der Forschungsdaten-Repositories in der Kristallographie	37
Tabelle 6: Anforderungen der Förderorganisationen in der Kristallographie	40
Tabelle 7: Auswertung der Forschungsdaten-Repositories in der Ökologie	44
Tabelle 8: Anforderungen der Förderorganisationen in der Ökologie	47
Tabelle 9: Auswertung der Forschungsdaten-Repositories in den Sozialwissenschaften	52
Tabelle 10: Anforderungen der Förderorganisationen in den Sozialwissenschaften	54

7. Abbildungsverzeichnis

Abbildung 1: Das Logo des Human Genome Project	11
Abbildung 2: Zugang zu vs. Bedeutung von Informationsressourcen	12
Abbildung 3: Die Datenpublikationspyramide	16

8. Abkürzungsverzeichnis

ABCD	Access to Biological Collections Data
ADS	Archaeology Data Service
AHAD	Australian Historical Archaeological Database
AHDS	Arts and Humanities Data Service
AHRC	Arts and Humanities Research Council
ALW	Aard-en Levenswetenschappen, Bereich der Geo- und Lebenswissenschaften der NWO
ANDS	Australian National Data Service
APS	American Physical Society
ARCS	Australian Research Collaboration Service
ASCII	American Standard Code for Information Interchange
ASSDA	Australian Social Science Data Archive
AUD	Australlian dollar
BADC	British Atmospheric Data Centre
BBSRC	Biotechnology and Biological Sciences Research Council
CAD	Computer Aided Design
CATI	Computer-Assisted Interviewing
CC	Creative Commons
CCDC	Cambridge Crystallographic Data Centre
CESSDA	Council of European Social Science Data Archives
CIDOC	
CRM	CIDOC Conceptual Reference Model
CIF	Crystallographic Information File; Crystallographic Information Framework
CNI	Coalition for Networked Information
COD	Crystallography Open Database
CODATA	Committee on Data for Science and Technology
CSD	Cambridge Structural Database
DAACS	Digital Archaeological Archive of Comparative Slavery
DAI	Deutsches Archäologisches Institut
DANS	Data Archiving and Networked Services
DCC	Digital Curation Centre
DDI	Data Documentation Initiative
DFG	Deutsche Forschungsgemeinschaft
DINI	Deutsche Initiative für Netzwerkinformation
DMP	Data Management Plan
DNA	deoxyribonucleic acid
DOI	Digital Object Identifier
DwC	Darwin Core
EASY	DANS Electronic Archiving System
EML	Ecological Metadata Language
EPSRC	Engineering and Physical Sciences Research Council
ESA	Ecological Society of America
ESDS	Economic and Social Data Service
ESRC	Economic and Social Research Council

ESS	European Social Survey
ESSD	Earth System Science Data
FP7	Seventh Framework Programme for Research
FWF	Fonds zur Förderung der wissenschaftlichen Forschung
GBIF	Global Biodiversity Information Facility
GDJ	Geoscience Data Journal
GEOSS	Global Earth Observation System of Systems
GESIS	chem. Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen
GIS	Geoinformationssystem
GRIB	Grid In Binary
HDF	Hierarchical Data Format
HGP	Human Genome Project
ICD	International Classification of Diseases
ICPSR	Interuniversity Consortium for Political and Social Research
ICSU	International Council for Science
IfA	Institute for Archaeologists
IFDO	International Federation of Data Organizations for the Social Sciences
InChI	International Chemical Identifier
IPCC	Intergovernmental Panel on Climate Change
ISO	International Organization for Standardisation
IUCr	International Union of Crystallography
JIF	Journal Impact Factor
JISC	Joint Information Systems Committee
JOAD	Journal of Open Archaeology Data
KII	Kommission Zukunft der Informationsstruktur
KNB	Knowledge Network for Biocomplexity
NACJD	National Archive of Criminal Justice Data
NCS	National X-ray Crystallography Service
NEH	National Endowment for Humanities
NERC	National Environment Research Council
NetCDF	Network Common Data Format
NHMRC	National Health and Medical Research Council
NIH	National Institutes of Health
NOAA	National Oceanic and Atmospheric Administration
NSF	National Science Foundation
NSF-CHE	NSF Division of Chemistry
NSF-SBE	NSF Social, Behavioral and Economic Sciences Directorate
NSIDC	National Snow and Ice Data Center
NWO	Niederländische Organisation für Wissenschaftliche Forschung
OAI-ORE	Open Archives Initiative Object Reuse and Exchange
OECD	Organisation for Economic Co-operation and Development
OJIMS	Overlay Journal Infrastructure for Meteorological Sciences
OpenDOAR	Directory of Open Access Repositories
PID	Persistent Identifier
PRC	Publishing Research Consortium
RatSWD	Rat für Sozial- und Wirtschaftsdaten
RDF	Resource Description Framework

RMetS	Royal Meteorological Society
RSC	Royal Society of Chemistry
SMILES	Simplified Molecular-Input Line-Entry System
SOM	Supplemental Online Material
SSHRC	Social Sciences and Humanities Research Council of Canada
STM	International Association of Scientific, Technical & Medical Publishers
tDAR	the Digital Archaeological Record
TDWG	chem. Taxonomic Databases Working Group
UKDA	UK Data Archive
UKPMC	UK PubMed Central
UKRIO	UK Research Integrity Office
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organization
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
URN	Uniform Resource Name
URN:NBN	URN for National Bibliography Numbers
WDC	World Data Center
WDCC	World Data Center for Climate
WMO	World Meteorological Organisation

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Masterarbeit zum Thema „Enhanced publications – Integration von Forschungsdaten beim wissenschaftlichen Publizieren“ selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle wörtlichen oder sinngemäßen Übernahmen aus anderen Quellen sind als solche gekennzeichnet.

Berlin, den 1. Oktober 2012



Dieses Werk steht unter einer [Creative Commons Namensnennung-Nicht-kommerziell 3.0
Deutschland Lizenz](https://creativecommons.org/licenses/by-nc/3.0/de/).